Eighteen Essays in Non-Euclidean Geometry

Editors:

Vincent Alberge
Department of Mathematics
Fordham University
441 East Fordham Road
Bronx, NY 10451
USA

Athanase Papadopoulos
Institut de Recherche Mathématique Avancée
CNRS et Université de Strasbourg
7 Rue René Descartes
67084 Strasbourg Cedex
France

2010 Mathematics Subject Classification: 03B30, 51A05, 51F15, 51F20, 51M04, 51M05, 51M09, 51M10, 51M16, 51M20, 51M25, 51N15, 52A15, 52A55, 52B15, 52C25, 05C62, 53-01, 53A05, 53A30, 53A35, 53C35, 53C45

Key words: Non-Euclidean geometry, spherical geometry, hyperbolic geometry, Busemann type geometry, curvature, geographical map, non-euclidean area, non-euclidean volume, Brahmagupta's formula, Ptolemy's theorem, Casey's theorem, Sforza's formula, Seidel's problem, infinitesimal rigidity, static rigidity, Pogorelov map, Maxwell–Cremona correspondence, exterior hyperbolic geometry, de Sitter geometry, non-Euclidean conics, bifocal properties, focus-directrix properties, pencils of conics, projective geometry, convexity, duality, transition, Hermitian trigonometry, complex projective trigonometry, shape invariant, metric plane projective-metric plane

ISBN 978-3-03719-196-5

The Swiss National Library lists this publication in The Swiss Book, the Swiss national bibliography, and the detailed bibliographic data are available on the Internet at http://www.helveticat.ch.

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in other ways, and storage in data banks. For any kind of use permission of the copyright owner must be obtained.

© 2019 European Mathematical Society

Contact address:

European Mathematical Society Publishing House
Seminar for Applied Mathematics
ETH-Zentrum SEW A21
CH-8092 Zürich
Switzerland

Email: info@ems-ph.org
Homepage: www.ems-ph.org

Typeset using the authors' TEX files: Marco Zunino, Savona, Italy
Printing and binding: Beltz Bad Langensalza GmbH, Bad Langensalza, Germany
∞ Printed on acid free paper
9 8 7 6 5 4 3 2 1

Foreword

This book consists of a series of essays on non-Euclidean geometry in a broad sense, including the classical geometries of constant curvature (spherical and hyperbolic), but also de Sitter, anti-de Sitter, co-Euclidean, co-Minkowski, Hermitian geometries, and some others. A few central themes emerge from these essays, namely, trigonometry, convexity, rigidity, area, volume, and the relation between non-Euclidean geometry and projective geometry, in particular the connection with conics, quadrics and quadratic forms. Some essays deal with very classical questions and others address problems that are at the heart of present day research, but all of them are concerned with fundamental topics.

The reader will see that the various essays contain many references to old mathematical texts. This is a result of the fact that geometry is a rich world which is built up on a socle of knowledge acquired through the effort of many generations. For the authors of this foreword, reading primary sources has always been a rewarding activity: it is in these original texts that one can find the most profound ideas that dominate modern research.

Although the topics considered in the various essays that constitute this volume are interconnected, the essays are completely self-contained. They vary in terms of length and difficulty. Sometimes, the same topic is treated in more than one of them, with more or less details, but the points of view are often different. The essays should be useful to researchers and students of non-Euclidean geometry. They are intended to be references for the various topics they present.

We acknowledge support from the U.S. National Science Foundation grants DMS 1107452, 1107263, 1107367 "RNMS: GEometric structures And Representation varieties" (the GEAR Network) and we warmly thank Marco Zunino for his excellent editorial work. Part of this work was done while the second editor was visiting the Yau Mathematical Sciences Center of Tsinghua University (Beijing).

Vincent Alberge (New York),
Athanase Papadopoulos (Strasbourg and Beijing),
December 2018

Introduction

Classically, the expression “non-Euclidean geometry” refers to the two geometries of non-zero constant curvature, namely, the spherical and the hyperbolic.[1] Spherical geometry was developed in Greek antiquity and it attained there a high degree of matureness, especially in the work of Menelaus of Alexandria (1^{st}–2^{nd} centuries A.D.).[2] Hyperbolic geometry is a nineteenth century achievement. It was discovered by Lobachevsky, Bolyai and Gauss, with a few anticipated results obtained in the eighteenth century, especially in the work of J. H. Lambert who developed a geometry (which, from his point of view, was hypothetical), in which all the Euclidean axioms hold except the parallel postulate and where the latter is replaced by its negation. Although Lambert’s goal was to find a contradiction in the consequences of this system of axioms (like many other prominent geometers, he thought that such a geometry cannot exist), the net result of his investigations is a collection of interesting theorems in hyperbolic geometry.[3]

In the present volume, the term “non-Euclidean geometry” is used in a broader sense, including geometries such as de Sitter, anti-de Sitter and others which can be developed, in analogy with the three classical geometries, in the framework of projective geometry, using the language of conics, quadrics and quadratic forms. Studying the three classical geometries in the setting of projective geometry is also a nineteenth century achievement; it is due to the visions of Cayley, Klein, Beltrami, Poincaré, and a few others.

Today, non-Euclidean geometry in this generalized sense is a very active research field and it seemed to us that editing a book containing self-contained surveys touching several aspects of this domain was desirable for researchers and those who want to learn this subject. This is the reason for which this book is published.

---

[1] The expression “non-Euclidean geometry,” to denote spherical and hyperbolic geometry, was coined by Gauss.

[2] The first English edition of Menelaus’ work just appeared in press: R. Rashed and A. Papadopoulos, *Menelaus’ Spherics*: *early translation and al-Māhānī/al-Harawī’s version,* Scientia Graeco-Arabica 21, Walter de Gruyter, Berlin, 2017.

[3] J. H. Lambert, “Theorie der Parallellinien,” in *Die Theorie der Parallellinien von Euklid bis auf Gauss, eine Urkundensammlung zur Vorgeschichte der nichteuklidischen Geometrie* (P. Stäckel and F. Engel, eds.), B. G. Teubner, Leipzig 1895. French translation by A. Papadopoulos and G. Théret, *La théorie des lignes parallèles de Johann Heinrich Lambert,* Sciences dans l’Histoire, Blanchard, Paris, 2014. There is no English translation available of this work.

Geometry is one of these mathematical fields (probably the unique one to such a great extent) where the texts of the ancient great authors keep their full value, and going through them is not a matter of vain curiosity nor a matter of being interested in history, but it is the best way to understand present day mathematics. In this sense, the rapid growth of non-Euclidean geometry in the last few decades, after its revival by Thurston in the 1970s, is also a return to the sources of geometry. This is why the names of Menelaus, Pappus, Euler, Lambert, Lagrange, Lexell, Fuss, Schubert, Chasles, Study, Story, and several others are mentioned at several occasions in the various essays that constitute this volume. Going throughout these works also shows how slow is the process of development of geometry, despite the rapid (quantitative) growth of the literature in this field. The slow evolution process of mathematics is marked out by abrupt changes attached to names such as those we just mentioned.

Before describing in some detail the content of this volume, we would like to make a few remarks on two topics that are at the heart of the various surveys that constitute it, namely, the notion of area in non-Euclidean geometry, and the geometry of conics.

In spherical geometry, the area of a triangle is (up to a constant multiplicative factor) the excess of its angle sum with respect to two right angles. In other words, the area of a triangle with angles A, B, C is equal to $A+B+C-\pi$, up to a constant factor that does not depend on the choice of the triangle. In hyperbolic geometry, the area of a triangle is (again up to a constant factor) the deficiency of the angle sum with respect to two right angles, that is, $\pi - (A + B + C)$. The fact that in spherical (respectively hyperbolic) geometry the angle excess (respectively the angle deficiency) of an arbitrary triangle is positive, is probably the most important feature in that geometry. As a matter of fact, there exist classical proofs of the trigonometric formulae in spherical and hyperbolic geometries that are exclusively based on these properties. The oldest such proofs that we are aware of are due to L. Gérard,[4] a student of Poincaré. Since the trigonometric formulae contain in essence all the geometric information on a space, spherical and hyperbolic geometry are essentially the geometries where angle excess or, respectively, angle deficiency of all triangles are positive. Thus, it is not surprising that several chapters in the present volume concern the notion of area and its use in non-Euclidean geometry.

After area, one naturally considers volume in higher-dimensional spherical and Euclidean geometries, and here we enter into the realm of difficult problems. One might recall in this respect that Gauss, in a letter to Wolfgang Bolyai (the father of János Bolyai, the co-discoverer of hyperbolic geometry), dated March 6, 1832,[5] after he gave an outline of a proof of the area formula for a hyperbolic triangle as angle deficiency, asked his friend to suggest to his son to work on the determination of volumes of tetrahedra in three-dimensional hyperbolic space. It is conceivable that Gauss was unaware of the fact that this subject was extensively studied by Lobachevsky during the same period, but he probably knew from his own experience that obtaining formulae for volumes of tetrahedra in hyperbolic and spherical spaces is a difficult problem. It turns out indeed that there is no simple formula for the volume of a hyperbolic (or spherical) tetrahedron in terms of the dihedral angles that is comparable to the formula for the area of a triangle. The existing formulae (first obtained by Lobachevsky) give expressions of volume in terms of an integral function which bears the name Lobachevsky function and which is closely connected to the Euler dilogarithm function. These formulae involve the value of the Lobachevsky function at the dihedral angles of the tetrahedron.

---

[4] L. Gérard, *Sur la géométrie non euclidienne,* Thèse № 768, Faculté des Sciences de l'Université de Paris, Gauthier-Villars, Paris, 1892.

[5] See Gauss's *Collected works*, Vol. 6, p. 221, and the article by P. Stäckel and F. Engel "Gauss, Die beiden Bolyai und die nichteuklidische Geometrie," *Math. Annalen* 2 (1897), 149–167.

Several computations of volumes of tetrahedra were conducted after Lobachevsky's work but the progress was slow. Some of the most important conjectures that are open in three-dimensional geometry of hyperbolic (or spherical) space concern volume. We mention incidentally that the computation of volumes of hyperbolic tetrahedra is the subject of Chapter 8 of Thurston's notes on three-dimensional geometry.[6]

Talking about volume, let us also mention that there are interesting formulae for volumes of Euclidean tetrahedra that are due to Euler. The latter, in a letter to his friend Christian Goldbach, dated November 14, 1750,[7] announced a formula for the volume of a simplex in terms of its side lengths, a three-dimensional analogue of Heron's formula for the area of a triangle.[8] Later on, he gave the proof of that formula, in a paper published in 1758.[9] In the same paper, he provided several other formulae for volumes of Euclidean simplices, including a formula in terms of side lengths, and another one in terms of lengths of the three edges abutting on a solid angle along with the three plane angles that these sides form.

Finally, let us note that computing area and volume, beyond those of polygons and polyhedra, is another subject that has a long history whose origin can be traced back to the work of Hippocrates of Chios (5th century B.C.) who studied areas of lunes (intersections of two discs) and other figures, and which was brought to a high degree of sophistication in the work of Archimedes (3rd century B.C.). The reader can imagine that computing areas and volumes without the use of the mechanical methods of modern calculus is always a challenging problem.

---

[6] W. P. Thurston, *Geometry and topology of three-manifolds*, Princeton Lectures Notes, Princeton, N.J., 1979.

[7] Letter № 149 in Euler's volume of his *Opera omnia* containing his correspondence with Goldbach, Series quarta A, Vol. IV, Birkhäuser, Basel, 2014.

[8] Heron's formula is contained in the *Codex costantinopolitanus Palatii Veteris,* ed. and transl. by E. M. Bruin, Leiden, 1964.

[9] L. Euler, Demonstratio nonnullarum insignium proprietatum, quibus solida hedris planis inclusa sunt praedita, *Novi Commentarii academiae scientiarum petropolitanae* 4 (1752), 140–160, in *Opera omnia,* Series 1, Vol. XXVI, 94–108.

Let us pass to our second topic, projective geometry and conics. We note right away that the study of conics dates back to Greek antiquity, where the most important treatise on the subject is the multi-volume work of Apollonius of Perga (3rd–2nd c. B.C.),[10] which remained until the seventeenth century one of the most fundamental and most read mathematical texts among mathematicians. A conic is the intersection of a plane with a right cone whose basis is circular. If we think of projective geometry as the study of properties of figures that are invariant by projections, it is in the study of conics that we find the root of this field: the cone vertex is the center of projection. The notion of cross-ratio as a projective invariant also dates back to Greek antiquity, it is used in Menelaus' *Spherics*.[2]

In his 1859 paper "A sixth memoir upon quantics,"[11] Cayley included the study of Euclidean and spherical geometries in the setting of projective geometry, where the ground space for each of these geometries is the interior of an appropriate conic in projective space, the conic itself becoming the "absolute" of the space. He noted that the distance function in such a geometry may be defined in terms of the logarithm of the cross-ratio. In the same paper, Cayley made his famous statement that "descriptive geometry is all geometry,"[12] an idea which was taken up by Klein later on, who, in his 1871 paper "Über die sogenannte Nicht-Euklidische Geometrie" (On the so-called non-Euclidean geometry),[13] included hyperbolic geometry as well in the picture, using different kinds of conics and quadrics in dimensions 2 and 3 respectively. In this work, Klein also gave formulae for the distance functions using the cross-ratio in the Euclidean and the two classical non-Euclidean geometries.

Among the various attempts to generalize hyperbolic geometry using the geometry of quadrics, we may mention a section in a paper of Poincaré which he wrote in 1887, called "Sur les hypothèses fondamentales de la géométrie" (On the fundamental hypotheses of geometry).[14] This paper is poorly known to mathematicians. The section which interests us here is concise, it contains ideas with no attempts for precise statements and proofs. We provide an English translation of it in the prologue of this volume, after the present introduction.

---

[10] The authoritative version is the one by R. Rashed, in 4 volumes: *Les Coniques,* Tome 1, Livre I, de Gruyter, Berlid 2008; Tome 2, Livre IV, ibid., 2009; Tome 3, Livre V, ibid., 2008; Tome 4, Livres VI et VII, ibid, 2009.

[11] A. Cayley, "A sixth memoir upon quantics," *Phil. Trans. R. Soc. Lond.* 149 (1859), 61–90. Reprinted in Vol. II of Cayley's *Collected mathematical papers.*

[12] "Descriptive geometry" is the name Cayley used for projective geometry.

[13] F. Klein, "Über die sogenannte Nicht-Euklidische Geometrie," Vorgelegt von A. Clebsch, *Nachricheten von der Kgl. Gesellschaft der Wissenschaften zu Göttingen,* № 17 (30 August 1871). French version, "Sur la géométrie dite non euclidienne," translated by J. Hoüel, *Bull. sci. math. et astr.* 2 (1871), 341–351. See also the commentary in N. A'Campo and A. Papadopoulos, "On Klein's *So-called non-Euclidean geometry,*" in *Sophus Lie and Felix Klein*: *the Erlangen program and its impact in mathematics and in physics* (L. Ji and A. Papadopoulos, eds.), EMS Publishing House, Zürich, 2015, 91–136.

[14] H. Poincaré, "Sur les hypothèses fondamentales de la géométrie." *Bull. Soc. Math. France* 15 (1887), 203–216.

Eduard Study, in a paper published in 1907[15] whose translation is contained in the present volume, included in the projective setting other geometries than the three classical geometries.

In the rest of this introduction, we will give a quick survey of the various essays that constitute the present volume. We have divided this collection of essays into three parts. The first part (Chapters 1–12) is the longest, and it is concerned with spherical and hyperbolic geometries. The second part (Chapters 13–16) deals with geometries defined in the setting of projective geometry. The third part (Chapters 17 and 18) concern two other geometries, namely, Hermitian geometry, that is, the geometry of complex projective spaces, and an axiomatic plane geometry which is termed "non-elliptic metric plane in which every segment has a midpoint."

The content of each chapter of this volume is now described in detail.

Part I. Spherical and hyperbolic geometries

Chapter 1, written by Norbert A'Campo and Athanase Papadopoulos, is a survey of classical material about area in spherical and hyperbolic geometry. A theorem of Albert Girard stating that the area of a spherical triangle is equal (up to a constant multiple) to its angle sum, as well as its analogue in hyperbolic geometry, are reviewed. A formula due to Euler for the area of a spherical triangle in terms of its side lengths, again with its analogue in hyperbolic geometry, are used in order to give an equality for the distance between the midpoints of two sides of a spherical (respectively hyperbolic) triangle in terms of the third side. These equalities are quantitative versions of the formula expressing the fact that the sphere (respectively hyperbolic plane) is positively (respectively negatively) curved in the sense of Busemann. The essay contains further results related to area in non-Euclidean geometry together with some historical comments.

Chapter 2, by Elena Frenkel and Weixu Su, is based on variational methods introduced by Euler in non-Euclidean geometry. The authors provide detailed proofs of hyperbolic analogues of spherical results obtained by Euler using these methods. This includes a derivation of the trigonometric formulae and an area formula for hyperbolic triangles in terms of their side lengths. Euler considered the use of variational methods in spherical geometry as an application of the techniques of the calculus of variations, a field of which he had laid the foundations.[16]

---

[15] E. Study, "Beiträge zur nichteuklidische Geometrie," I.–III., *Amer. J. Math.* 29 (1907), 101–167.

[16] Cf. L. Euler, "Principes de la trigonométrie sphérique tirés de la méthode des plus grands et des plus petits," *Mémoires de l'Académie royale des sciences et belles-lettres* 9 (1755), 223–257, in *Opera omnia,* Series 1, Vol. XXVI, 277–308.

In Chapter 3, Elena Frenkel and Vincent Alberge study the hyperbolic analogue of a problem in spherical geometry solved by Friedrich Theodor von Schubert, a young geometer who became a successor of Euler at the Saint Petersburg Academy of Sciences and who worked on spherical geometry. Schubert's results include the determination of the loci of the vertices of triangles satisfying some given conditions, in the spirit of problems solved by Euler and his other young collaborators Anders Johan Lexell and Paul Heinrich Fuss. In this chapter, the authors give a solution in the hyperbolic setting of a problem solved by Schubert on the sphere, namely, to find, for a triangle with prescribed base and whose vertex varies on a given hypercycle consisting of points equidistant to the line containing the base, the point(s) for which the area of this triangle is maximal or minimal. They provide two different solutions of this problem.

The reader may note that in the Euclidean case, if we fix two vertices of a triangle and if we look for a family of triangles having the same area, the locus of the third vertex is a straight line parallel to the base of the triangle (in fact, in the Euclidean plane, hypercycles are straight lines). The situations in the hyperbolic and spherical cases are different from the Euclidean case since in these cases, the area of a triangle is not determined by a side and the corresponding altitude. The desired locus in these cases is not a curve equidistant to the basis of the triangle. It is interesting to note in this respect that Herbert Busemann solved in 1947 the following related problem: "To characterize the geometries in which the following property, which is satisfied in Euclidean geometry, holds: *The area of a triangle ABC depends only on the length of BC and the distance from A to the segment BC*." Busemann found that the only geometries that satisfy this property are what he called the Minkowski planes (that is, the metric spaces underlying the 2-dimensional normed vector spaces) in which the perpendicularity relation is symmetric (for an appropriate notion of perpendicularity).[17]

Chapter 4, by Himalaya Senapati, concerns medians in non-Euclidean geometry. A result due to J. H. Lambert, from his *Theorie der Parallellinien* (Theory of parallel lines), written in 1766,[3] says that for any equilateral triangle ABC with medians AA', BB' and CC' intersecting at O, we have $OA' = \frac{1}{3}AA'$, $OA' > \frac{1}{3}AA'$ and $OA' < \frac{1}{3}AA'$ in Euclidean, spherical and hyperbolic geometry respectively. Senapati shows, using non-Euclidean trigonometry, that Lambert's inequalities hold for arbitrary triangles. He also presents a collinearity result in spherical geometry, as an application of Menelaus' theorem.

---

[17] Cf. Theorem 50.9 in H. Busemann, *The geometry of geodesics,* Academic Press, New York, 1955, and H. Busemann, "Two-dimensional geometries with elementary areas," *Bull. Amer. Math. Soc.* 53 (1947), 402–407, reprinted in H. Busemann, *Selected works* (A. Papadopoulos, ed.), Vol. 1, Springer Verlag, Cham, 2018, 379–384.

Chapter 5 is also due to Himalaya Senapati. It concerns the construction in spherical geometry of a triangle whose three vertices lie on a given circle and whose three sides produced pass through three given points. Such a construction, in the Euclidean case, is a theorem in Pappus' *Collection*,[18] in the special case where the three points are aligned. Euler generalized the problem to the case where the three points are not necessarily aligned and he gave a construction in this general case in the Euclidean as well as in the spherical setting.[19] In the same volume in which Euler's solution appeared, his young collaborator Nicolaus Fuss[20] published another proof of the same construction for the Euclidean case.[21] Lagrange found a new short proof for the Euclidean case, reducing the problem to finding roots of quadratic equations that can be solved using ruler and compass.[22] In Senapati's essay, Lagrange's solution is adapted to the non-Euclidean setting.

The notion of constructibility by ruler and compass is at the root of geometry. It underwent several transformations, from the Greeks to Galois, passing through Descartes, Gauss, and others. Senapati's construction on the sphere, in Chapter 5, based on the Euclidean Lagrange solution of the generalized problem of Pappus, is in the tradition of Descartes, that is, reducing the construction problem to solutions of quadratic equations.

We note incidentally that the first time Descartes introduced the xy coordinates (what we call the "Cartesian coordinates") is in Book I of his *Géométrie*, precisely in the solution of a construction problem of Pappus. We also recall that the title of this book is "Des problèmes qu'on peut construire sans y employer que des cercles et des lignes droites" that is, "On problems that can be constructed using only circles and straight lines."

Chapter 6 is again due to Himalaya Senapati. It is based on two propositions of Menelaus' *Spherics* that are comparison results for angles in a triangle cut by a geodesic arc joining the midpoints of two sides. The two propositions, in Menelaus' treatise, belong to a group of propositions in which the author proves for spherical triangles the property called today the Busemann property for positively curved metric spaces. Senapati gives an improved version of the angle comparison result in the case of spherical triangles as well as a version for hyperbolic triangles.

---

[18] Pappus d'Alexandrie, *La collection mathématique,* Œuvre traduite pour la premiere fois du grec en français par P. Ver Eecke, Desclée de Brouwer, Paris and Bruges, 1933, Propositions 105, 107, 108, and 117 of Book VII.

[19] L. Euler, "Problematis cuiusdam Pappi Alexandrini constructio," *Acta Academiae scientarum imperialis petropolitanae*, Pars 1 (1780), 91–96. In *Opera omnia,* Series 1, Vol. XXVI, 237–242.

[20] Nicolaus Fuss (1755–1826) was initially Euler's secretary, and he became later his student, collaborator, and colleague at the Russian Academy of Sciences, and eventually the husband of his granddaughter Albertine.

[21] N. Fuss, "Solutio problematis geometrici Pappi Alexandrini," *Acta Academiae scientarum imperialis petropolitanae*, Pars 1 (1780), 97–104.

[22] J.-L. Lagrange, "Solution algébrique d'un problème de géométrie," in *Oeuvres de Lagrange*, Vol. IV, Gauthier-Villars, Paris 1868, 335–339. There is a confusion about the authorship of this work. It first appeared in a memoir under the name of J. de Castillon (Giovanni Francesco Mauro Melchiorre Salvemini, also called il Castiglione, after his birthplace, Castiglione del Valdarno), in which the latter declares that this proof is due to Lagrange, and after that, he provides his own proof, which is different from Lagrange's. The part of Castillon's memoir that corresponds to Lagrange's proof appeared eventually in Lagrange's *Œuvres* under his own name.

In Chapter 7, Dmitriy Slutskiy gives a proof of the trigonometric formulae of spherical and hyperbolic geometries using a kinematic method, following the work of Joseph Marie de Tilly (1837–1906). This approach is explained in a memoir by the latter, published in 1870 under the title *Études de mécanique abstraite* (Studies in abstract mechanics).[23] The basic tools used are two functions that de Tilly calls $\mathrm{eq}(r)$ and $\mathrm{circ}(r)$ respectively, the first one being the length of a curve at distance r from a geodesic of length one, and the second one being the length of a circle of radius r. The work is model-free.

Chapter 8, by Son Lam Ho, is concerned with the classical Gauss–Bonnet formula for surfaces of variable curvature. This is again a result on area. It says that the integral of the Gaussian curvature over a closed surface in the Euclidean space $\mathbb{R}^3$, with respect to the area form, is equal to 2π times the Euler characteristic of the surface. This formula is one of the first formulae that establish a relation between, on the one hand, topology (the Euler characteristic) and on the other hand, geometry (curvature and area). The author explains the relation between the Gauss–Bonnet formula and the formula for the area of a triangle in terms of angle excess.

Chapter 9, by Charalampos Charitos and Ioannis Papadoperakis, is based on a theorem of Euler on cartography. Euler's result says that there is no perfect map from an open subset of the sphere into the Euclidean plane. Here, a perfect map is a smooth map that preserves distances infinitesimally along the meridians and parallels as well as angles between these lines. The theorem is generally quoted with a different (and wrong) statement in the literature by historians of mathematics (and sometimes mathematicians who followed the historians' writings) who misunderstood the statement.[24] In the present essay, Euler's precise statement and a detailed proof of it are provided.

---

[23] J.-M. de Tilly, *Études de mécanique abstraite,* Mémoires couronnée et autres mémoires publiés par l'Académie Royale de Belgique, Vol. XXI, 1870.

[24] An instance of a wrong quote is contained in R. Ossermann's "Mathematical mapping from Mercator to the millenium," in *Mathematical adventures for students and amateurs* (D. F. Hayes and T. Shubin, eds.), The Mathematical Society of America, Washington (D.C.), 2004, 237–257, Spectrum. Theorem 1 there, attributed to Euler, states the following: *It is impossible to make an exact scale map of any part of a spherical surface*, an exact scale map meaning, in that paper, a map that preserves distances up to scale. The author refers to Euler's paper "De repraesentatione superficiei sphaericae super plano published" in *Acta Academiae scientarum imperialis petropolitanae,* Pars 1 (1777), 107–132 (*Opera omnia,* Series 1, Vol. XXVIII, 248–275). The theorem stated in Ossermann's paper is not what Euler proves. This result was known long before Euler, since Greek antiquity, and Euler's theorem is much more involved than that. The same error occurs in the paper "Curvature and the notion of space" by A. Knoebel, J. Lodder, R. Laubenbache, and D. Pengelley, in *Mathematical masterpieces,* Springer Verlag, New York, 2007, 159–227, and in many other papers.

Chapter 10, by Charalampos Charitos, is based again on certain projection maps from (subsets of) the sphere to the Euclidean plane that were considered by Euler. The study concerns more precisely the class of area-preserving projections of class C^2 that satisfy the further property that they send all meridians and parallels to perpendicular curves in the Euclidean plane. This class of maps was highlighted by Euler in his memoir "De repraesentatione superficiei sphaericae super plano" (On the representation of spherical surfaces onto the plane).[24] The distortion from conformality of these projections is compared with that of the so-called "Lambert cylindrical equal area projection," a map introduced by Lambert and rediscovered by Euler. The author shows that the distortion of the Lambert projection has remarkable extremal properties in the class considered.

Thus, Chapters 9 and 10 concern the search of maps between non-Euclidean and Euclidean geometries that are best in an appropriate sense. It is also interesting to see that by their study of such maps, Euler, Lambert and the geographers that preceded them are in fact predecessors of the modern theory of quasiconformal mappings. The interested reader may refer to the recent survey entitled "Quasiconformal mappings, from Ptolemy's geography to the work of Teichmüller."[25]

Chapter 11, by Nikolai Abrosimov and Alexander Mednykh, consists of two parts. The first one concerns area, and the second one volume, in spherical and hyperbolic geometries.

In the first part, the authors survey various classical formulae by Euler, Cagnoli, Lhuillier and others on the area of a triangle in spherical geometry and their counterparts in hyperbolic geometry. They also present non-Euclidean analogues of a Euclidean area formula due to Bretschneider and other recent formulae for the area of non-Euclidean triangles and quadrilaterals, together with non-Euclidean versions and generalizations of an identity of Ptolemy concerning cyclic quadrilaterals, and related identities due to Casey, and others.

In the second part, the authors survey classical formulae for volumes of various kinds of polyhedra due to Schläfli, Lobachevsky, Bolyai, Coxeter, and Vinberg. They discuss in particular the case of an orthoscheme, that is, a simplex in which the edges are mutually orthogonal. A three-dimensional orthoscheme has three right dihedral angles, the other dihedral angles being termed *esesential*. Schläfli provided a formula for the volume of a spherical orthoscheme in terms of its essential dihedral angles, using a function which is now called the Schläfli function. Lobachevsky and Bolyai obtained formulae for volumes of hyperbolic orthoschemes. A theory of volume of ideal hyperbolic tetrahedra originated in the work of Lobachevsky done in the second quarter of the nineteenth century. This theory was revived by Thurston and Milnor in the late 1970s. Vinberg considered in detail the case of tetrahedra having at least one vertex at infinity.

---

[25] A. Papadopoulos, "Quasiconformal mappings, from Ptolemy's geography to the work of Teichmüller," in *Uniformization, Riemann–Hilbert correspondence, Calabi–Yau manifolds, and Picard–Fuchs equations* (L. Ji and S.-T. Yau, eds.), Advanced Lectures in Mathematics 42, Higher Education Press, Beijing, and International Press, Boston, 2018, pp. 237–315.

Together with these developments, the authors in Chapter 11 present several other relatively recent results on volume, including a result of Abrosimov on a problem of Seidel asking for the expression of the volume of an ideal hyperbolic tetrahedron in terms of the determinant and the permanent of its so-called “Gram matrix.” They also report on a formula, due to Derevin and Mednykh, expressing the volume of a compact hyperbolic tetrahedron in terms of its dihedral angles, and on a formula due to Sforza concerning the volume of a compact tetrahedron in hyperbolic or spherical 3-space. They also present formulae for volumes of spherical and hyperbolic octahedra with various kinds of symmetries.

Chapter 12 by Ivan Izmestiev is concerned with rigidity problems of bar-and-joint frameworks in Euclidean, spherical and hyperbolic geometries. A bar-and-joint framework is an object made of rigid bars connected at their ends by universal joints, that is, joints that allow the incident bars to rotate in any direction. The questions that are addressed in this chapter generalize rigidity questions for polyhedra. The author approaches them from two points of view, which he calls the static and the kinematic. In the scientific jargon, statics is the science of equilibrium of forces, and kinematics the science of motion. In the present setting, the static rigidity of a framework refers to the fact that every system of forces with zero sum and zero moment applied on it can be compensated by stresses in the bars, whereas its kinematic rigidity refers to the absence of deformations that keep the lengths of bars constant to the first order. The author shows that these two notions of rigidity of a framework are equivalent. More generally, he proves that a framework in a Euclidean, spherical, or hyperbolic space has the same number of static and kinematic degrees of freedom. Here, the number of static degrees of freedom is the dimension of the vector space of unresolvable loads, that is, of systems of forces applied to the nodes that cannot be compensated by stresses in the bars. In particular, a framework is statically rigid if the number of its static degrees of freedom is zero. The number of kinematic degrees of freedom is the dimension of the vector space of non-trivial infinitesimal isometric deformations, that is, deformations that are not induced by an isometry of the ambient space.

The subject of statics, in the sense used in this essay, finds its roots in the 19th century works of Poinsot, Möbius, Grassmann, and others. Questions of infinitesimal rigidity of smooth surfaces were addressed in the 20th century by H. Weyl, A. D. Alexandrov, and A. V. Pogorelov. Izmestiev in Chapter 12 provides a projective interpretation of statics which allows him to prove a projective invariance property of infinitesimal rigidity and to establish a correspondence between the infinitesimal motion of a Euclidean framework and its geodesic realization in the two geometries of constant nonzero curvature. He refers to the fact that the number of degrees of freedom of a Euclidean framework is a projective invariant as the Darboux–Sauer correspondence. He also reviews the so-called “Maxwell–Cremona” correspondence for a framework in Euclidean and in the two non-Euclidean geometries. This correspondence establishes an equivalence between the existence of a self-stress of a framework (a collection of stresses in its bars that resolves a zero load), a reciprocal diagram (a framework whose combinatorics is dual to the combinatorics of a given framework), and a so-called “polyhedral lift" (which, in the Euclidean case, is a vertical lift to 3-space which has the property that the images of the vertices of every face are coplanar).

Part II. Projective geometries

Chapters 13 and 14 inaugurate the second part of the volume. They consist of a translation (made by Annette A’Campo-Neuen) and a short commentary (by Annette A’Campo-Neuen and Athanase Papadopoulos) on a paper by Eduard Study, published in 1907, entitled “Beiträge zur Nicht-Euklidischen Geometrie. I.”[15] In this paper, Study introduces what he calls the *exterior plane hyperbolic geometry*, which turns out to be the geometry we call today the de Sitter geometry. This is a geometry of the complement of hyperbolic space, when this space is realized as the Cayley–Klein disc model sitting in the projective plane. The exterior space is also the space of lines in the hyperbolic plane. A distance function (which is not a metric in the sense we intend it today) is defined on pairs of points in this exterior space, and the way the distance between two points is defined depends on the position of these points, more precisely, on whether the line joining them intersects the unit circle, or is tangent to it, or is disjoint from it. Study gives a characterization of triangles for which the triangle inequality holds, and of triangles for which the reverse triangle inequality (called the time inequality) holds. He also discusses the cases where the distance between two points is equal to the length of the longest curve joining them.

It is interesting to see that the ideas expressed by Study in this paper are explored in modern research. They appear in the two essays that follow (Chapters 15 and 16).

Chapter 15 is also due to Ivan Izmestiev. It consists of an exposition of the theory of conics in spherical and hyperbolic spaces. The author presents this theory in full detail, including classifications of conics from the algebraic and analytic points of view, the theory of pencils, the characteristic properties of foci, axes, and centers, the bifocal properties of spherical conics, the various concepts of polarity and duality as well as the non-Euclidean versions of several results including Poncelet, Brianchon, Pascal and Chasles’ theorems, and Ivory’s lemma. The latter states that the diagonals of a quadrilateral formed by four confocal conics have equal lengths. De Sitter space is also included in the picture.

The topic of spherical conics is classical. Back in 1788, Nicolaus Fuss wrote a memoir on spherical conics, entitled “De proprietatibus quibusdam ellipses in superficie sphaerica descriptae” (On some properties of an ellipse traced on a spherical surface).[26] Michel Chasles made a systematic study of spherical conics in his 1860 article “Résumé d’une théorie des coniques sphériques homofocales et des surfaces du second ordre homofocales” (Summary of a theory of spherical homofocal conics and second-order homofocal surfaces).[27] William Story, in a paper on non-Euclidean geometry published in 1881,[28] computed areas of conics in the hyperbolic plane; his results are expressed in terms of elliptic integrals. In 1883, he wrote a paper on hyperbolic conics entitled “On non-Euclidean properties of conics.”[29] Today, the subject of non-Euclidean conics is almost forgotten, and even, the study of Euclidean conics is no more part of the curricula. Izmestiev’s essay, based on the articles by Chasles and Story, revives this subject, showing at the same time that Euclidean and non-Euclidean geometries (including de Sitter) can be approached via the study of (Euclidean) conics. From this point of view, a non-Euclidean conic is represented by a pair of quadratic forms on Euclidean space, the first one (assumed to be non-degenerate) representing the absolute of a non-Euclidean geometry, and the second one being the conic in the non-Euclidean space realized by the first conic.

In Chapter 16, François Fillastre and Andrea Seppi start from the fact, which we mentioned several times, that in dimension two, a model for each of the three classical geometries can be obtained as the interior of (a connected component of the complement of) a certain conic, the latter becoming the “absolute” of the space. The authors survey several two- and three-dimensional geometries that may be obtained, in the tradition of Klein, in the setting of projective geometry. These geometries include, besides the classical Euclidean, spherical and hyperbolic geometries, de Sitter and anti-de Sitter geometries, as well as geometries that the authors call co-Euclidean and co-Minkowski. The last two are respectively the geometry of the space of hyperplanes in Euclidean space and that of space-like hyperplanes in Minkowski geometry. Some of these geometries appear as dual to or as limits of other geometries. The developments of the theory presented in this essay use duality theory in Euclidean space associated to convex sets (polar duality). A convex complementary component of a quadric in $\mathbb{R}^3$ appears as a model of hyperbolic space whereas the other component appears as a model for de Sitter space and is also described as the space of hyperplanes in hyperbolic space. The definition of anti-de Sitter space, as a Lorentzian manifold of curvature -1, uses bilinear forms of signature 2.

---

[26] N. Fuss, “De proprietatibus quibusdam ellipses in superficie sphaerica descriptae,” *Nova acta Academiae scientiarum imperialis petropolitanae* 3 (1785), 90–99.

[27] M. Chasles, “Résumé d’une théorie des coniques sphériques homofocales et des surfaces du second ordre homofocales,” *J. Math. Pures Appl.* (1860), 425–454.

[28] W. E. Story, “On the non-Euclidean trigonometry,” *Amer. J. Math.* 4 (1881), 332–335.

[29] W. E. Story, “On non-Euclidean properties of conics,” *Amer. J. Math.* 5 (1883), 358–382.

Questions regarding transitions between geometries are also addressed in Chapter 16, and in this setting, elliptic and de Sitter geometries limit to co-Euclidean geometry. Likewise, Minkowski and co-Minkowski geometries appear as limit geometries of classical ones. The authors also discuss connections and volume forms for their model spaces, leading to connections and volume forms for the degenerate co-Euclidean and co-Minkowski geometries. A map they call "infinitesimal Pogorelov map" is introduced, between the hyperbolic and Euclidean spaces, and between anti-de Sitter and Minkowski spaces. They then study embedded (especially convex) surfaces in the 3-dimensional model spaces, as well as geometric transitions of surfaces. They show that the notion of curvature transits between various geometries, under rescaling in co-Euclidean and co-Minkowski geometries.

Part III. Other geometries

In Chapter 17, Boumediene Et-Taoui provides a survey of Hermitian trigonometry, that is, trigonometry in complex projective space. In particular, he presents the two laws of sines and the law of cosines for a triangle in the complex projective plane. The subject of Hermitian trigonometry is classical. It was first studied by Blaschke and Terheggen in their paper "Trigonometria Hermitiana."[30] The exposition in Chapter 17 follows in part that of Ulrich Brehm in his article "The shape invariant of triangles and trigonometry in two point homogeneous spaces."[31] Et-Taoui defines at the same time the notion of polar triangle associated to a given triangle in complex projective space and he uses it to derive the second law of sines. He also establishes relations between invariants introduced by Brehm and others. He notes that the spherical geometry developed by Eduard Study is a special case of the Hermitian trigonometry developed in this chapter. In particular, the trigonometric formulae of spherical trigonometry can easily be deduced from the Hermitian trigonometric formulae. The essay also contains remarks on trigonometry in general symmetric spaces, a subject where almost all the basic questions are still open.

In Chapter 18, Victor Pambuccian presents a result on triangles of fixed area in a geometry he describes as the "Bachmann non-elliptic metric plane in which every pair of points has a midpoint." This geometry is defined in an axiomatic way in terms of groups, involutions in these groups and axioms concerning the relations among these involutions. The expression "elliptic metric plane" refers to a geometry in which there are three line reflections whose composition is the identity. The result presented says that for any given pair of points A and C and for any point B varying in the plane in such a way that the area of the triangle ABC is constant, the locus of the midpoints of AB and CB consists of two lines symmetric with respect to AC. This result is in the trend of a theorem attributed to Lexell (with various proofs by Lexell, Euler, Legendre, Steiner, Lebesgue, and several others)[32] concerning the locus of the third vertex of a spherical triangle with two given vertices and fixed area.

---

[30] W. Blaschke and H. Terheggen, "Trigonometria Hermitiana," *Rend. Sem. Mat. Univ. Roma Ser.* 4 (1939), 153–161.

[31] U. Brehm, "The shape invariant of triangles and trigonometry in two point homogeneous spaces," *Geom. Dedicata* 33 (1990), 59–76.

This set of essays should give the reader a taste of the fundamental notions of non-Euclidean geometry and an idea of a variety of problems that are studied in this very rich branch of mathematics.

Vincent Alberge (New York),
Athanase Papadopoulos (Strasbourg and Beijing),
December 2018

---

[32] A. J. Lexell, "Solutio problematis geometrici ex doctrina sphaericorum," *Acta Academiae scientarum imperialis petropolitanae*, Pars 1 (1781), 112–126; L. Euler, "Variae speculationes super area triangulorum sphaericorum," *Nova acta Academiae scientarum imperialis petropolitanae* 10 (1792), 47–62, in *Opera omnia,* Series 1, Vol. XXIX, 253–266; A.-M. Legendre, *Éléments de géométrie* (11th ed.). Firmin Didot, Paris 1817; J. Steiner, "Sur le maximum et le minimum des figures dans le plan, sur la sphère et dans l'espace général," *J. Math. Pures Appl.* 6 (1841), 105–170; H. Lebesgue, "Démonstration du théorème de Lexell," *Nouvelles annales de mathématiques* 14 (1855), 24–26.

Prologue
An excerpt from “Sur les hypothèses fondamentales de la géométrie”[*]

Henri Poincaré

We already know three two-dimensional geometries:

1. Euclidean geometry, where the angle sum in a triangle is equal to two right angles;
2. Riemann’s geometry, where this sum is greater than two right angles;
3. Lobachevsky’s geometry, where it is smaller than two right angles.

These three geometries rely on the same fundamental hypotheses, if we except Euclid’s *postulatum* which the first one admits and the two others reject. Furthermore, the principle according to which two points completely determine a line entails an exception in Riemann’s geometry and entails no exception in the other two cases.

When we restrict ourselves to two dimensions, Riemann’s geometry allows a very simple interpretation; it does not differ, as is well known, from spherical geometry, provided we decide to give the name *lines* to the great circles of the sphere.

I shall start by generalizing this interpretation in such a way that it can be extended to Lobachevsky’s geometry.

Consider an arbitrary order-two surface.[1] We agree to call *lines* the diametral plane[2] sections of this surface, and *circles* the non-diametral plane sections.

---

[*] H. Poincaré “Sur les hypothèses fondamentales de la géométrie,” *Bull. Soc. Math. France* 15 (1887) 203–216. Translation by A. Papadopoulos.

[1] [Translator’s note] A degree-two algebraic surface, also called a quadric.

[2] [Translator’s note] For a given quadric, the locus of midpoints of the system of chords that have a fixed direction is a plane, called a *diametral plane* of the quadric. In the case where the quadric has a center of symmetry, a diametral plane is a plane passing through this center of symmetry. There are more general points of symmetry for a quadric. In analogy with the case of the great circles and the small circles of a sphere, Poincaré calls a *line* the intersection of a quadric surface with a diametral plane, and a *circle* the intersection of a quadric with an arbitrary plane.

It remains to define what we mean by the angle made by two lines which intersect and by the length of a line segment.

Through a point on the surface, let two diametral plane sections pass (which we decided to call *lines*). Let us then consider the tangents to these two plane sections and the two rectilinear generatrices of the surface which pass through the given point. These four lines (in the usual sense of the word) have a certain anharmonic ratio.[3] The angle that we want to define will then be the logarithm of the anharmonic ratio if the two generatrices are real, that is, if the surface is a one-sheeted hyperboloid. In the other case, our angle will be the same logarithm divided by $\sqrt{-1}$.

Let us consider an arc of conic which is part of a diametral plane section (this is what we agreed to call a *line segment*). The two endpoints of the arc and the two points at infinity of the conic have, as any system of four points situated on a conic, a certain anharmonic ratio. Then we agree to call *length* of the given segment the logarithm of this ratio if the conic is a hyperbola and this same logarithm divided by $\sqrt{-1}$ if the conic is an ellipse.

There will be, between the angles and the lengths thus defined, a certain number of relations, and these will constitute a collection of theorems which are analogous to those of plane geometry.

This collection of theorems can then take the name *quadratic geometry*, since our starting point was the consideration of a quadric or a second order fundamental surface.

There are several quadratic geometries, because there are several kinds of second order surfaces.

If the fundamental surface is an ellipsoid, then the quadratic geometry does not differ from Riemann's geometry.[4]

If the fundamental surface is a two-sheeted hyperboloid, then the quadratic geometry does not differ from that of Lobachevsky.

If this surface is an elliptic paraboloid, then the quadratic geometry is reduced to that of Euclid; this is a limiting case of the preceding two.

It is clear that we have not exhausted the list of quadratic geometries; because we have not considered neither the one-sheeted hyperboloid nor its numerous degeneracies.

Thus, we can say that there are three main quadratic geometries, which correspond to the three types of second-order surfaces with center.

On the other hand, we have to add to them the geometries that correspond to the limiting cases and Euclidean geometry will be part of them.

---

[3] [Translator's note] The term *anharmonic ratio* means the cross ratio.

[4] [Translator's note] In the 19th century, the term *Riemann's geometry* was commonly used for the geometry of the sphere.

Why did the geometry of the one-sheeted hyperboloid, up to now, escape the theorists? This is because it implies the following propositions.

1. The distance between two points situated on the same rectilinear generator of the fundamental surface is zero.
2. There are two kinds of lines, the first one corresponding to the elliptic diametral sections and the other one to the hyperbolic diametral sections. It is impossible, using any real motion, to make a line of the first kind coincide with a line of the second.
3. It is impossible to make a line coincide with itself by a real rotation around one of its points, as it happens in Euclidean geometry when we rotate a line by 180° around one of its points.

All the geometers have implicitly assumed that these three propositions are false, and indeed these three propositions are so much contrary to the habits of our spirit that its was not possible to the founders of our geometries to make such a hypothesis and think of stating it.

Contents

Foreword v
Vincent Alberge and Athanase Papadopoulos

Introduction vii
Vincent Alberge and Athanase Papadopoulos

Prologue xxi
An excerpt from "Sur les hypothèses fondamentales de la géométrie"
Henri Poincaré (*translated by Athanase Papadopoulos*)

PART I. Spherical and hyperbolic geometries

Chapter 1
Area in non-Euclidean geometry 3
Norbert A'Campo and Athanase Papadopoulos

Chapter 2
The area formula for hyperbolic triangles 27
Elena Frenkel and Weixu Su

Chapter 3
On a problem of Schubert in hyperbolic geometry 47
Vincent Alberge and Elena Frenkel

Chapter 4
On a theorem of Lambert:
medians in spherical and hyperbolic geometries 57
Himalaya Senapati

Chapter 5
Inscribing a triangle in a circle in spherical geometry 67
Himalaya Senapati

Chapter 6
Monotonicity in spherical and hyperbolic triangles 81
Himalaya Senapati

Chapter 7
De Tilly's mechanical view on hyperbolic and spherical geometries 93
Dmitriy Slutskiy

Chapter 8
The Gauss–Bonnet theorem and the geometry of surfaces 113
Son Lam Ho

Chapter 9
On the non-existence of a perfect map from the 2-sphere to the Euclidean plane 125
Charalampos Charitos and Ioannis Papadoperakis

Chapter 10
Area preserving maps from the sphere to the Euclidean plane 135
Charalampos Charitos

Chapter 11
Area and volume in non-Euclidean geometry 151
Nikolay Abrosimov and Alexander Mednykh

Chapter 12
Statics and kinematics of frameworks in Euclidean and non-Euclidean geometry 191
Ivan Izmestiev

PART II. Projective geometries

Chapter 13
Contributions to non-Euclidean geometry I 237
Eduard Study (translated by Annette A'Campo-Neuen)

Chapter 14
Notes on Eduard Study's paper "Contributions to non-Euclidean geometry I" 253
Annette A'Campo-Neuen and Athanase Papadopoulos

Chapter 15
Spherical and hyperbolic conics 263
Ivan Izmestiev

Chapter 16
Spherical, hyperbolic, and other projective geometries: convexity, duality, transitions 321
François Fillastre and Andrea Seppi

PART III. Projective geometries

Chapter 17
Hermitian trigonometry 413
Boumediene Et-Taoui

Chapter 18
A theorem on equiareal triangles with a fixed base 427
Victor Pambuccian

List of contributors 439

PART I

Spherical and hyperbolic geometries

Chapter 1

Area in non-Euclidean geometry

Norbert A'Campo and Athanase Papadopoulos

Contents

1. Introduction 3
2. Angle excess and Girard's theorem in spherical geometry 4
3. The hyperbolic analogue 9
4. Area and side length 18
References 21

1 Introduction

We discuss various appearances of the notion of area of triangles in non-Euclidean geometry (spherical and hyperbolic). We start by recalling the theorem of Girard which gives the area of a spherical triangle in terms of its angle sum, and its classical proof based on the notion of lune. We give an analogous proof for the corresponding theorem in hyperbolic geometry, after introducing an object which replaces lunes in this setting, namely, a triangle having two vertices at infinity. We then recall a formula for Euler for the area of a spherical triangle in terms of its side lengths, and its analogue for a hyperbolic triangle. The two formulae are used in order to give a precise formula for the distance between the midpoints of two sides of a spherical (respectively hyperbolic) triangle, in terms of the third side. These formulae are precise quantitative versions of the positivity (respectively negativity) of the curvature in the sense of Busemann, in the cases of the sphere and of the hyperbolic plane.

At several places in this chapter, we make relations between classical theorems and the work of Herbert Busemann. An edition of his Collected works was published recently by Springer Verlag, see [9].

2 Angle excess and Girard's theorem in spherical geometry

Let us start by recalling the notion of lune in spherical geometry.

Two great circles on the sphere intersect at two antipodal points. The complement of these two great circles has four components, see Figure 1. A *lune* is one such component. In other words, a lune is a portion of the sphere bounded by two semicircles (that is, half of great circles). Each lune has two vertices, and the angles at these two vertices are equal. We recall that on the sphere, an angle made by two arcs of great circles meeting at a point is the dihedral angle made by the two planes in 3-space containing the two great circles. Two triangles are called *colunar* if they share a common edge and if their union forms a lune (Figure 2). This notion was known and extensively used by Menelaus of Alexandria (1$^{\text{st}}$–2$^{\text{nd}}$ c. BC) in his *Spherics*, a major work (probably the most important one ever written) on spherical geometry. An edition of the *Spherics*, with and English translation, appeared recently [28].

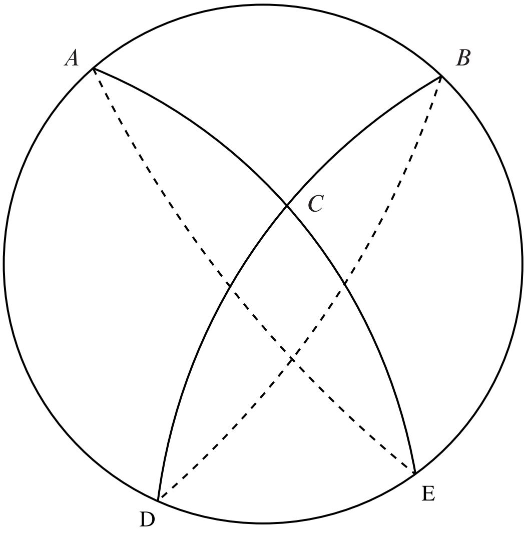


Figure 1. The complement of two great circles consists in four lunes

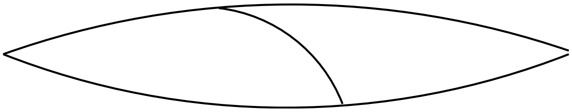


Figure 2. Two colunar triangles

Given a spherical triangle with vertices A, B, C, its *angle excess* is the quantity $A + B + C - \pi$. (We denote by the same letter a vertex and the angle made by the two sides meeting at that vertex.) We give a short proof of the fact that the angle excess of every spherical triangle is positive. This result is classical: it was known to Menelaus; it is Proposition 12 of his *Spherics* [28]. Menelaus' proof of this theorem is also reproduced in [24]. In the case of a hyperbolic triangle ABC, one defines the *angle deficit* as the quantity $\pi - (A + B + C)$. In the next theorem, we give a different proof of Menelaus' result and we prove at the same time that in hyperbolic geometry, the angle deficit of an arbitrary triangle is positive.

Theorem 2.1. *In spherical* (*respectively hyperbolic*) *geometry, the angle excess* (*respectively deficit*) *of an arbitrary triangle is positive.*

Proof. The proof in both cases is by contradiction.

Consider first the case of spherical geometry. We know that there exists a triangle whose angle excess is positive (take for instance a triangle whose three angles are right angles). Assume that there exists a triangle whose angle excess is non-positive. Then, since the angle excess function on triangles is continuous, moving continuously the three vertices of the first triangle to the vertices of the second one, we find a triangle whose angle excess is zero. But with such a triangle we can tile the whole plane, and it is easy to see that a plane tiled by triangles whose angle sum is equal to two right angles is Euclidean. This is a contradiction. We conclude that the angle excess of an arbitrary triangle is positive.

In hyperbolic geometry, we start with a triangle whose angle deficit is positive. For this, we may either consider an ideal triangle, whose three angles are zero, or we may take a triangle whose three vertices are close to infinity, that is, they are close to the vertices of an ideal triangle. The rest of the proof is analogous. □

We now turn to Girard's theorem and its proof.

We let $\mathcal{K}$ be one eighth of the total area of the sphere. In other words, $\mathcal{K}$ is the area of a triangle of angles $\frac{\pi}{2}, \frac{\pi}{2}, \frac{\pi}{2}$. We consider areas of lunes. Two lunes having equal angles have the same area, since they are isometric. From this, it follows easily that the area of a lune of angle A is equal to $\frac{4\mathcal{K}}{\pi}A$.

We have the following:

Theorem 2.2 (Girard's theorem). *For any spherical triangle ABC, we have*

$$\mathrm{Area}(ABC) = \frac{2\mathcal{K}}{\pi}(A + B + C - \pi).$$

Proof. The proof is based on computation of areas of lunes. We consider the figure formed by the union of the three great circles that contain the edges of the triangle ABC, see Figure 3. We have

$$\begin{aligned}\text{Area}(ABDCA) &= \frac{4\mathcal{K}}{\pi}A,\\ \text{Area}(BCEAB) &= \frac{4\mathcal{K}}{\pi}B,\\ \text{Area}(CAFBC) &= \frac{4\mathcal{K}}{\pi}C.\end{aligned}$$

Adding both sides of these three equations, we get

$$\begin{aligned}&\text{Area}(ABDCA) + \text{Area}(BCEAB) + \text{Area}(CAFBC)\\ &\quad= \text{Area}(AFDEA) + \text{Area}(BCEAB) + \text{Area}(CAFBC)\\ &\quad= \text{Area}(ABC) + \text{Area}(EFD) + 4\mathcal{K}\\ &\quad= 2\,\text{Area}(ABC) + 4\mathcal{K},\end{aligned}$$

which gives

$$2\,\text{Area}(ABC) + 4\mathcal{K} = \frac{4\mathcal{K}}{\pi}A + \frac{4\mathcal{K}}{\pi}B + \frac{4\mathcal{K}}{\pi}C$$

where the term $4\mathcal{K}$ on the left-hand side represents the area of the heimisphere. ☐

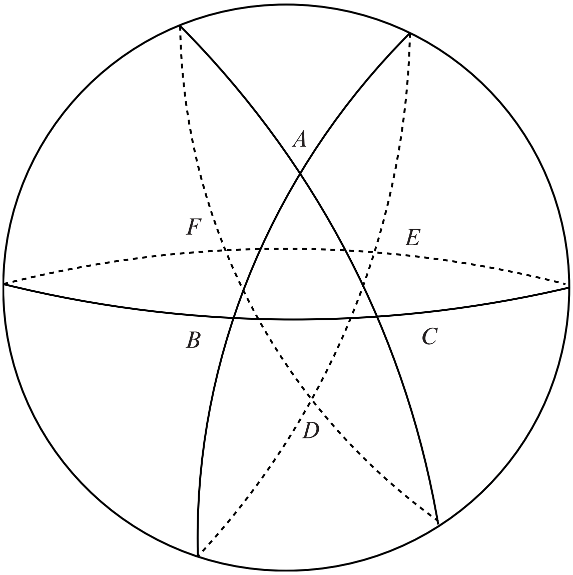


Figure 3

Notes. 1.— We used the fact that the area of a lune is proportional to its angle. Strictly speaking, this needs a proof, similar to the one that Euclid gave for Proposition 1 of Book VI of the *Elements* saying that triangles and parallelograms having the same base and the same hight have equal areas. A similar remark holds for the hyperbolic theorem (Theorem 3.2 below).

2.— Theorem 2.2 is attributed to Albert Girard (1595–1632), a French mathematician from Lorraine, who studied and worked in the Netherlands [30]. He stated this theorem in his *Invention nouvelle en l'algèbre* (New invention in algebra) [16] (1629), in an appendix called "De la mesure de la superficie des triangles et polygones sphériques, nouvellement inventée" (On the measure and surface of spherical triangles and polygons, recently discovered).

A proof of the same theorem is attributed to Thomas Harriot (1560–1621), but it seems that no written proof by him survives, and therefore this remains hypothetical. We refer the reader to the discussion of this fact in the work [2], p. 28ff.

Lagrange, in his memoir "Solution de quelques problèmes relatifs aux triangles sphériques avec une analyse complète de ces triangles" (Solution of some problems relative to spherical triangles with a complete analysis of these triangles) [17], considers that the proof that Girard gave is not satisfying and he writes that the theorem should rather be attributed to Bonaventura Francesco Cavalieri (1598–1647), who gave a complete proof of it in his *Directorium generale uranometricum* (A general guide for celestial measures), Bologna, 1632. Lagrange also refers to Cavalieri's "beautiful proof reported on by Wallis" (1616–1703). This proof is contained in Vol. II of Wallis' *Collected works* [32], p. 875; cf. Figure 4.

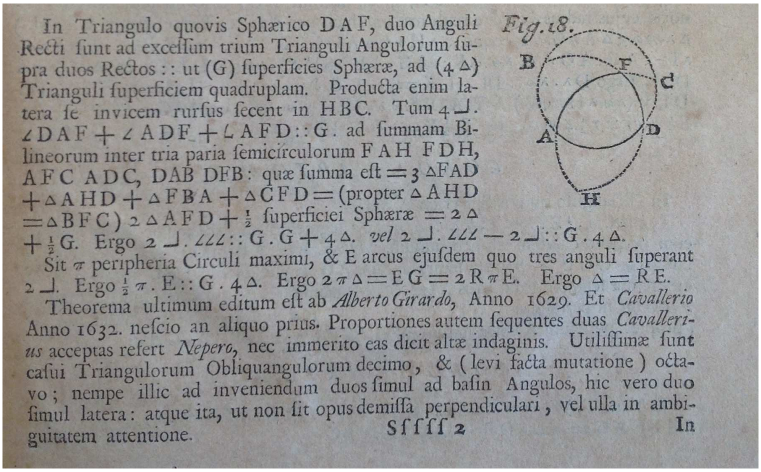


Fig. 18.
B
F
C
A
D
H

In Triangulo quovis Sphærico DAF, duo Anguli Recti ſunt ad exceſſum trium Trianguli Angulorum ſupra duos Rectos :: ut (G) ſuperficies Sphæræ, ad ($4\triangle$) Trianguli ſuperficiem quadruplam. Producta enim latera ſe invicem rurſus ſecent in HBC. Tum $4\lrcorner . \angle DAF + \angle ADF + \angle AFD :: G$. ad ſummam Bilineorum inter tria paria ſemicirculorum FAH FDH, AFC ADC, DAB DFB: quæ ſumma eſt $= 3\triangle FAD + \triangle AHD + \triangle FBA + \triangle CFD =$ (propter $\triangle AHD = \triangle BFC$) $2\triangle AFD + \frac{1}{2}$ ſuperficiei Sphæræ $= 2\triangle + \frac{1}{2}G$. Ergo $2\lrcorner . \angle\angle\angle :: G . G + 4\triangle$. *vel* $2\lrcorner . \angle\angle\angle - 2\lrcorner :: G . 4\triangle$.

Sit π peripheria Circuli maximi, & E arcus ejuſdem quo tres anguli ſuperant $2\lrcorner$. Ergo $\frac{1}{2}\pi . E :: G . 4\triangle$. Ergo $2\pi\triangle = EG = 2R\pi E$. Ergo $\triangle = RE$.

Theorema ultimum editum eſt ab *Alberto Girardo*, Anno 1629. Et *Cavallerio* Anno 1632. neſcio an aliquo prius. Proportiones autem ſequentes duas *Cavallerius* acceptas refert *Nepero*, nec immerito eas dicit altæ indaginis. Utiliſſimæ ſunt caſui Triangulorum Obliquangulorum decimo, & (levi facta mutatione) octavo; nempe illic ad inveniendum duos ſimul ad baſin Angulos, hic vero duo ſimul latera: atque ita, ut non ſit opus demiſſa perpendiculari, vel ulla in ambiguitatem attentione.

Sſſſſ 2 In

Figure 4. From Wallis' proof of Girard's Theorem ([32], Vol. 2, p. 875)

Euler, in his memoir [11], gives a proof of Theorem 2.2 using the calculus of variations. In his memoir [12], he gives a proof using lunes. Figure 5 is extracted from Euler's paper [12]. In the picture in the middle, the triangle ABC is represented with its three colunar triangles. The proof uses the fact that the three points a, b, c, which are the antipodals of the vertices of A, B, C of the initial triangle, are at equal mutual distances than the points A, B, C (Euler writes: "even though in the figure they seem much different"). Euler's proof then consists in computing the areas of the eight triangles in the figure, $ABC, AbC, BcA, CaB, abC, bcA, caB, abc$ and adding them. The sum is the total area of the sphere, and grouping terms gives the result. We shall mention below another result of Euler on spherical area, contained in his memoir [13]. The interested reader is referred to the book [1] for an English translation and a commentary of several memoirs of Euler on spherical trigonometry.

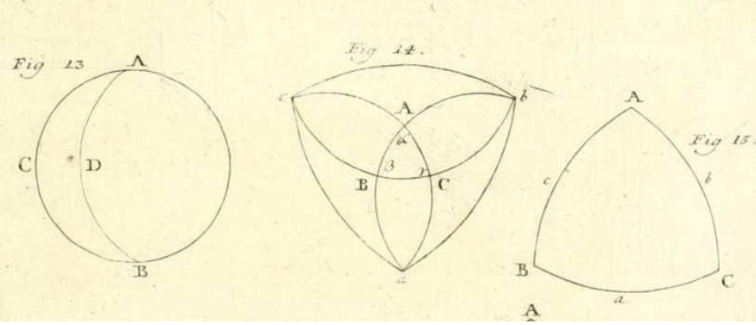


Figure 5. From Euler's paper [12]

Prouhet, in his paper [26] writes that "one must add that Girard does not claim that his proof is rigorous, but only as a provisional probability, promising to give a rigorous proof." He adds: "It is evident that Euler did not read Girard's work. The proof that is attributed to him is due to Cavalieri," and he gives the reference [10] Prouhet, in the same paper, describes the content of Cavalieri's book, and he ends his review with the following biographical note: "Cavalieri (B.) was born in Milan in 1598. He entered the Jesuates order and dedicated himself to mathematics. He became the devoted student of Galileo. He was distracted from his studies by violent gout attacks and he died on December 3, 1667" (the correct year should be 1647).

3.— Cavalieri's name, mentioned in the preceding note, is tightly related to area and volume in Euclidean space, since he is the author of the so-called *Cavalieri principle* which says that two bodies in Euclidean space have the same volume if the areas of their sections by planes which are parallel to a given plane are equal. For obvious reasons, his result is considered as a precursor of integral calculus. Note that Cavalieri obtained this result before the invention of the differential calculus of

Newton and Leibniz. We also mention a famous related problem from the modern period, relating volume and areas of sections (and, more generally, n-volume and $(n-1)$-volume in $\mathbb{R}^n$), called the Busemann–Petty problem (this is Problem 1 in the problem list [7]; Petty was a student of Busemann). The problem asks the following. Let K_1 and K_2 be two convex bodies in $\mathbb{R}^n$ that are symmetric with respect to the origin and such that

$$\mathrm{Vol}_{n-1}(K_1 \cap H) \leq \mathrm{Vol}_{n-1}(K_2 \cap H)$$

for any hyperplane H passing through the origin. Is it true that $\mathrm{Vol}_n(K_1) \leq \mathrm{Vol}_n(K_2)$?

The problem resisted several years before a complete solution was obtained in 1999, but there are several variants and generalizations of this problem which constitute a living research topic.

3 The hyperbolic analogue

We give a proof of a hyperbolic analogue of Girard's formula. The proof is the one contained in Thurston's book [31], p. 83ff. and it is close in spirit to a proof of Gauss of the same theorem, which we recall after Thurston's proof.

We start with a hyperbolic triangle $S(\alpha)$ with angles $\alpha, 0, 0$. Thus, two of its vertices are at infinity; cf. Figure 6. The isometry type of such a triangle is determined by the angle α. Let $A(\alpha)$ be the area of $S(\alpha)$.

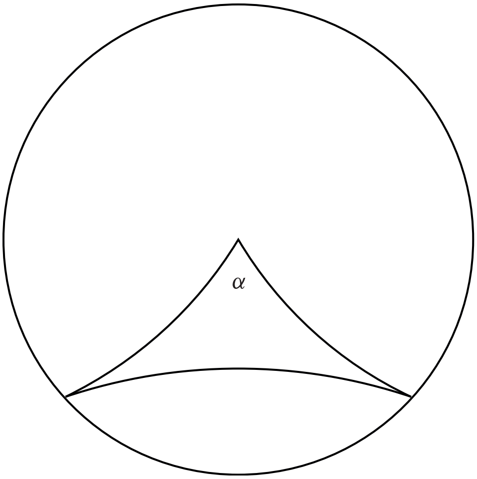


Figure 6

The following lemma is used in the proof of the main result (Theorem 3.2) and is also a special case of it.

Lemma 3.1. *We have* $A(\alpha) = \frac{P}{\pi}(\pi - \alpha)$, *where* P *is the area of an ideal triangle.*

Proof. From Figure 7, we have

$$A(\alpha + \beta) = A(\alpha) + A(\beta) - A(0),$$

and $A(0) = P$, that is, the area of an ideal triangle. (We recall that all ideal triangles are congruent, since the group of motions of the hyperbolic plane acts transitively on triples of distinct points at infinity; therefore, any two such triangles have the same area).

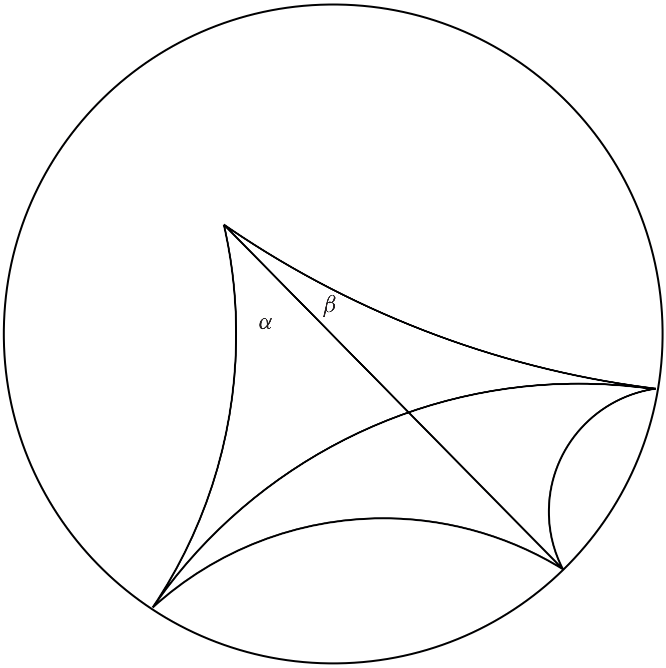


Figure 7

Thus, we have

$$P - A(\alpha + \beta) = P - A(\alpha) + P - A(\beta).$$

Therefore the function

$$\alpha \longmapsto P - A(\alpha)$$

is linear in an appropriate sense.[1]

---

[1] The function is not defined on a vector space, but the meaning of the word linear used here is clear.

Thus, we can set $P - A(\alpha) = k\alpha$ where k is some constant.

For $\alpha = \pi$, the triangle $S(\alpha)$ is degenerate, and its area is zero. Thus, we have $A(\pi) = 0$.

From $k\pi = P - A(\pi)$, we get $P = k\pi$.

This gives $A(\alpha) = k(\pi - \alpha) = \frac{P}{\pi}(\pi - \alpha)$ for all α. ☐

From this, we can easily deduce the following well-known formula for the area of a hyperbolic triangle:

Theorem 3.2. *We have*

$$\text{Area}(ABC) = \frac{P}{\pi}(\pi - A - B - C)$$

where P is, as above, the area of an ideal triangle.

Proof. Consider an ideal quadrilateral Q (a quadrilateral with four vertices at infinity). The two diagonals divide Q into four triangles having a common vertex and each of which has its two other vertices at infinity.

We call *lune* the union of two sectors sharing two vertical angles (Figure 8). These two sectors are isometric by the central symmetry whose centre is the intersection of the two diagonals.

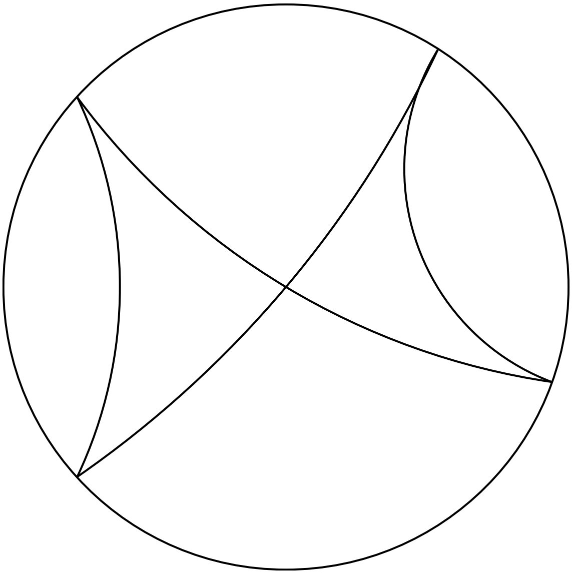


Figure 8

The area of a lune depends only on its central angle. By Lemma 3.1, the area of a lune with central angle α is equal to $2k(\pi-\alpha)$ with $k = P/\pi$, where P is, as above, the area of an ideal triangle.

Consider now a hyperbolic triangle ABC. We draw the three lunes whose central angles are the three angles of the triangle. The union of these three lunes is an ideal hexagon H (see Figure 9).

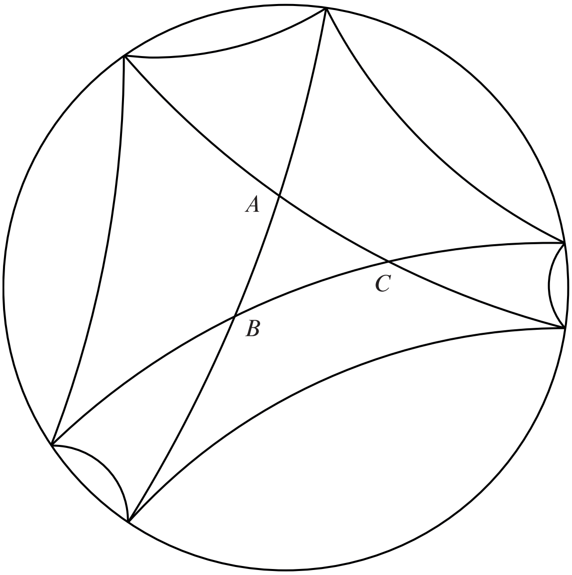


Figure 9

The area of the ideal hexagon H is 4 times the area of an ideal triangle, that is,

$$\text{Area}(H) = 4k\pi.$$

The sum of the areas of the three lunes is equal to the sum of the area of the hexagon H with two times the area of the triangle ABC. Thus, we have

$$2k(\pi - A) + 2k(\pi - B) + 2k(\pi - C) = 4k\pi + 2\,\text{Area}(ABC).$$

Grouping terms, we get

$$\text{Area}(ABC) = k(\pi - A - B - C). \qquad \square$$

Historical notes. To the best of our knowledge, the first proof of Theorem 3.2 is due to the Alsatian mathematician Johann Heinrich Lambert (1728–1777). It is contained in his *Theorie der Parallellinien* (Theory of parallels), written in 1766. This memoir is now available in French translation; see [22], an edition of this memoir together with a historical and a mathematical commentary. See also [23] and the review [27]; they both give an idea of the work of Lambert. In his memoir, Lambert developed the bases of hyperbolic geometry, without being sure that such a geometry exists. In fact, Lambert, like all his predecessors, thought that such a geometry cannot exist. His aim in developing a geometry where all the Euclidean postulates hold except the parallel postulate, and where this postulate fails, was to find a contradiction. In doing so, he proved several theorems of hyperbolic geometry, and unlike his predecessors, none of the counter-intuitive statements that he found concerning that geometry seemed to him to be a contradiction. Lambert's poof of Theorem 3.2 is contained in §81 of his memoir. We recall it here. It is amazingly simple. Lambert makes the hypothesis that there exists a triangle whose angle sum is less than two right angles. (He also shows that this is equivalent to the fact that any triangle satisfies this property.) He then shows that the angle excess function, that is, the function that assigns to a triangle with angles A, B, C the value $\pi - (A + B + C)$, is invariant under subdivision. To see this, consider a triangle ABC divided into two smaller triangles with angles A, E, F and B, C, D as in Figure 10. The angle deficits of the two smaller triangles are $\pi - (A + E + F)$ and $\pi - (B + C + D)$ respectively. Adding the two quantities, and using the fact that $D + E = \pi$, we find $2\pi - (A + E + F + B + C + D) = 2\pi - (A + B + C + F)$, which is the angle deficit of the triangle ABC. Lambert concludes that the angle deficit of a hyperbolic triangle is an additive magnitude; therefore it is proportional to area. In fact, instead of dividing a triangle into two triangles as we did, Lambert reasons on a triangle ABC divided into three triangles by three lines EF, FG, GE, as in Figure 11. The proof is similar. He continues the discussion of area in §82 and 83 of his memoir, noting that a similar formula for area, that is, Girard's formula, is valid in spherical geometry, with the angle deficit of triangles replaced by angle excess. He also discusses areas of polygons.

After Lambert, we turn to Gauss. His notes and correspondence on non-Euclidean geometry (spherical and hyperbolic) are contained in Volume VIII of his *Collected works* [15]. In a letter he wrote to Wolfgang Bolyai on March 6, 1832, Gauss gives the formula for the area of a hyperbolic triangle in terms of its angle deficit. See [15], Vol. VIII, p. 221. We recall the proof. It uses the notion of parallelism in hyperbolic geometry.

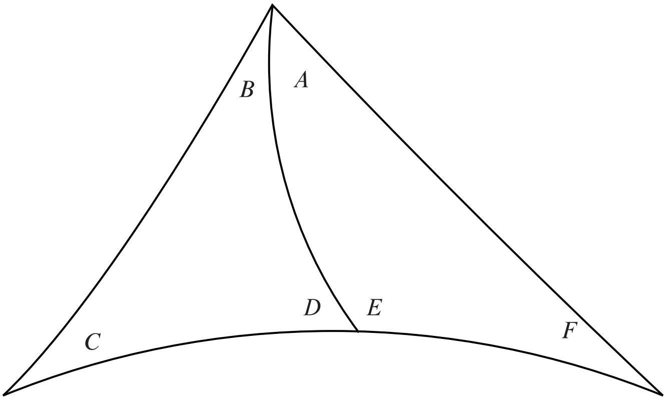


Figure 10. A triangle divided into two triangles

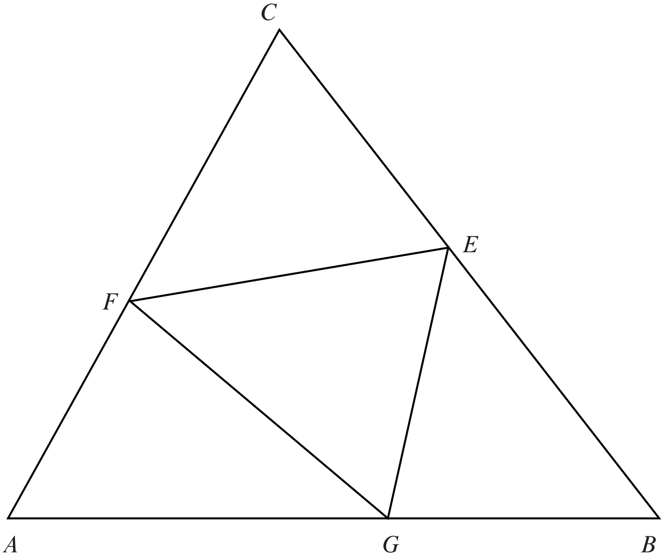


Figure 11. Lambert's proof of the area as angle deficit: additivity of angle deficit

In the hyperbolic plane, if L is a line and x a point not on L, a line *parallel* to L through x is a line that does not intersect L and that is the limit of two families of lines: one family of lines that intersect L and a second family of lines that do not intersect L. (See Figure 12.) This definition was used by the three founders of hyperbolic geometry. Lobachevsky, in his *Geometrische Untersuchungen zur Theorie der Parallellinien* (Geometric investigations on the theory of parallel lines) [18], studied in detail this relation of parallelism. The trigonometric formulae that he obtained involve this function.

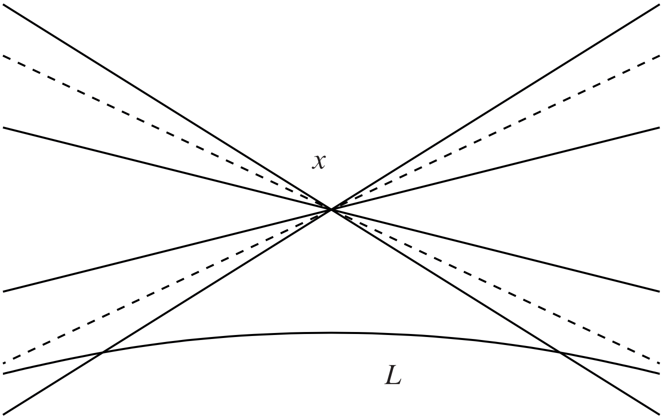


Figure 12. The two shaded lines are the parallels through x to the line L

Given a line L and a point P not on L, consider the two parallels to L from P and let ϕ the angle they make. Gauss knew that the area $A(\phi)$ bounded by L and these two parallels is finite (Figure 13). He also knew that the area of an ideal triangle is a constant (that is, this area does not depend on the choice of an ideal triangle). Let C be this value.

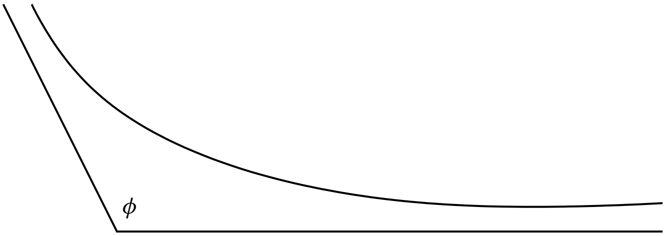


Figure 13. The area bounded by the given line and the two parallel rays which make an angle ϕ is denoted by $A(\phi)$

Consider the function

$$f(\phi) = A(\pi - \phi).$$

From Figure 14, we have, for any angle ϕ in $(0, \pi)$,

$$f(\pi - \phi) + f(\phi) = C.$$

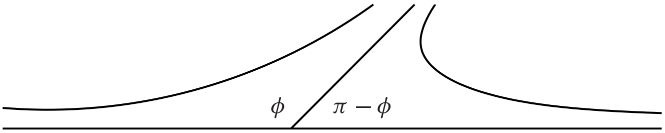


Figure 14

Consider now a division of the total angle at some point into three angles and let their values be $\pi - \phi, \pi - \psi$ and $\phi + \psi$ (Figure 15). We have

$$f(\phi) + f(\psi) + f(\pi - \phi - \psi) = C.$$

This gives

$$f(\phi) + f(\psi) = C - f(\pi - \phi - \psi) = f(\phi + \psi).$$

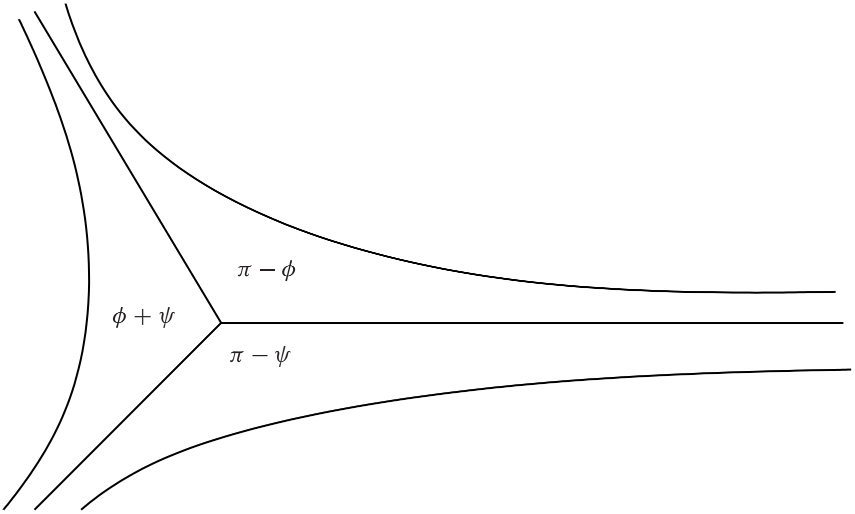


Figure 15

From this, Gauss deduces that the function f is linear, that is, it has the form

$$f(\phi) = \lambda\phi$$

where λ is a constant.

Now given an arbitrary triangle Δ, we produce three of its sides, as indicated in Figure 16.

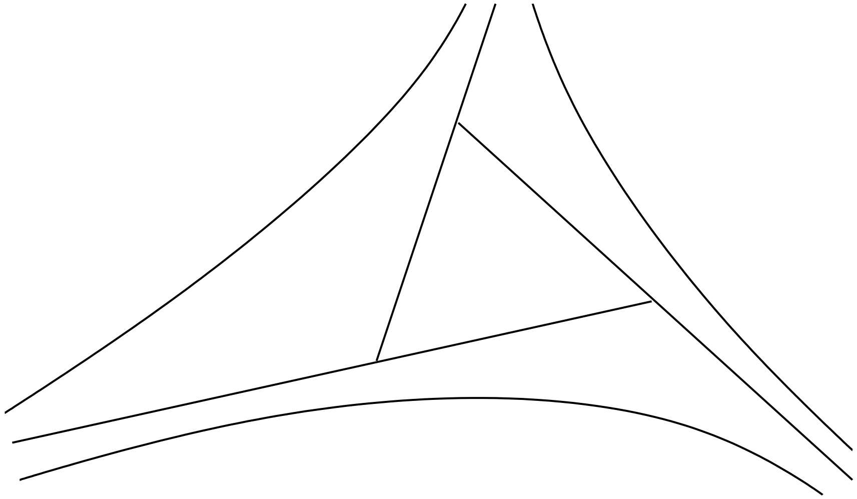


Figure 16

We obtain

$$C = f(\alpha) + f(\beta) + f(\gamma) + \text{Area}(\Delta),$$

which gives

$$\text{Area}(\Delta) = \lambda(\pi - \alpha - \beta - \gamma).$$

This completes the proof.

Gauss's proof is contained in the article [29] by Staeckel and Engel which reproduces the correspondence between Gauss and the two Bolyai.

Lobachevsky, in his *Pangeometry* gives another proof of the formula for the area of a hyperbolic triangle as angle deficit. See [21], p. 48. In the same treatise (p. 65), he gives a formula for the area of a hyperbolic triangle in terms of its side lengths, more precisely, in terms of the so-called angle of parallelism function of these sides. As a matter of fact, the computation of areas and volumes of non-Euclidean objects constitute a substantial part of Lobachevsky's treatise [21] and of some other works, e.g. his *Application of imaginary geometry to certain integrals* (1836) [19] and his "Géométrie imaginaire" (Imaginary geometry) (1837) [20].

4 Area and side length

In this section, ABC is a triangle (which may be spherical or hyperbolic) with side lengths a, b, c opposite to the vertices A, B, C, and Δ is its area. We shall consider that the sphere has constant curvature $+1$ and that the hyperbolic plane has constant curvature -1. In particular, the area formula (Theorems 2.2 and 3.2) become respectively

$$\Delta = A + B + C - \pi$$

and

$$\Delta = \pi - (A + B + C).$$

In his memoir [13], Euler gave the following formula expressing the area of ABC in terms of its side lengths.

Theorem 4.1 (Euler's formula for the area of a spherical triangle [13]). *The following identity holds*:

$$\cos\left(\frac{1}{2}\Delta\right) = \frac{1 + \cos(a) + \cos(b) + \cos(c)}{4\cos\left(\frac{1}{2}a\right)\cos\left(\frac{1}{2}b\right)\cos\left(\frac{1}{2}c\right)}.$$

Euler gave a first proof of this formula using a computation, and another proof using the calculus of variations. We shall also use the following analogue of Euler's formula for the case of a hyperbolic triangle, a proof based on the calculus of variations, given in the paper [14] in the present volume:

Theorem 4.2 (Euler's formula for area, hyperbolic analogue). *The following identity holds*:

$$\cos\left(\frac{1}{2}\Delta\right) = \frac{1 + \cosh(a) + \cosh(b) + \cosh(c)}{4\cosh\left(\frac{1}{2}a\right)\cosh\left(\frac{1}{2}b\right)\cosh\left(\frac{1}{2}c\right)}.$$

The following result, in the spherical case, is contained in [25] p. 115.

Theorem 4.3. *In a triangle ABC of area Δ, let D and E be the midpoints of AB and BC (Figure 17). We have, in the spherical case,*

$$\cos(DE) = \cos\left(\frac{1}{2}\Delta\right)\cos\left(\frac{1}{2}AC\right),$$

and in the hyperbolic case,

$$\cosh(DE) = \cos\left(\frac{1}{2}\Delta\right)\cosh\left(\frac{1}{2}AC\right),$$

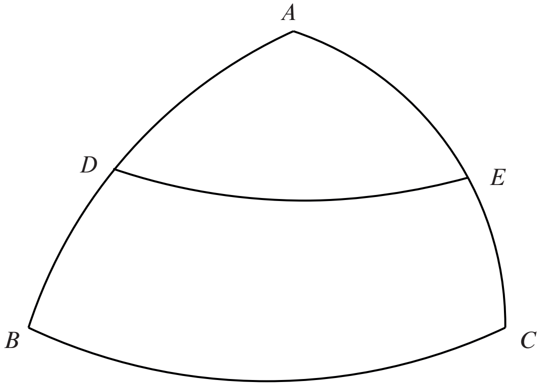


Figure 17. D and E are the midpoints of AB and AC

Proof. We provide the proof in the spherical case. In the hyperbolic case, the proof is analogous. The cosine law in the triangles ABC and ADE gives

$$\cos(A) = \frac{\cos(a) - \cos(b)\cos(c)}{\sin(b)\sin(c)} \tag{1}$$

and

$$\cos(A) = \frac{\cos(DE) - \cos\left(\frac{1}{2}b\right)\cos\left(\frac{1}{2}c\right)}{\sin\left(\frac{1}{2}b\right)\sin\left(\frac{1}{2}c\right)}. \tag{2}$$

Equating the two values of $\cos A$, we obtain

$$\cos(DE) = \frac{1}{\sin(b)\sin(c)}\Big((\cos(a) - \cos(b)\cos(c))\Big(\sin\left(\frac{1}{2}b\right)\sin\left(\frac{1}{2}c\right)\Big) + \cos\left(\frac{1}{2}b\right)\cos\left(\frac{1}{2}c\right)\sin(b)\sin(c)\Big).$$

Replacing in the right-hand side $\cos(b)$ and $\cos(c)$ by their values in terms of $\sin\left(\frac{1}{2}b\right)$, $\sin\left(\frac{1}{2}c\right)$, $\cos\left(\frac{1}{2}b\right)$, and $\cos\left(\frac{1}{2}c\right)$, and after some cancellations, we get

$$\begin{aligned}\cos(DE) &= \frac{\cos(a) + 2\cos^2\left(\frac{1}{2}b\right) + 2\cos^2\left(\frac{1}{2}c\right) - 1}{4\cos\left(\frac{1}{2}b\right)\cos\left(\frac{1}{2}c\right)}\\ &= \frac{\cos(a) + (1+\cos(b)) + (1+\cos(c)) - 1}{4\cos\left(\frac{1}{2}b\right)\cos\left(\frac{1}{2}c\right)}\\ &= \frac{1 + \cos(a) + \cos(b) + \cos(c)}{4\cos\left(\frac{1}{2}b\right)\cos\left(\frac{1}{2}c\right)}\\ &= \frac{1 + \cos(a) + \cos(b) + \cos(c)}{4\cos\left(\frac{1}{2}b\right)\cos\left(\frac{1}{2}c\right)\cos\left(\frac{1}{2}a\right)} \cdot \cos\left(\frac{1}{2}a\right).\end{aligned}$$

Using Euler's formula (Theorem 4.1), we obtain

$$\cos(DE) = \cos\left(\frac{1}{2}\Delta\right)\cos\left(\frac{1}{2}a\right),$$

which is the required formula. □

Corollary 4.4. *Consider a family of spherical or hyperbolic triangles ABC having fixed area and a fixed side length, say, the length of AC is fixed. Then, the distance between the midpoints of the sides AB and BC is also constant, for all triangles in the family.*

This result is attributed by Papelier ([25], p. 115), in the spherical case, to Gudermann.

Corollary 4.5. *Using the notation of Theorem 4.5, we have, in the spherical case,*

$$DE > \frac{1}{2}AC, \tag{3}$$

and in the hyperbolic case,

$$DE < \frac{1}{2}AC. \tag{4}$$

Notes. 1.— The result of inequation (3) in Corollary 4.5 concerning the sphere was known to Menelaus; this is Proposition 27 of his *Spherics* [28].

2.— The cosine laws (formulae (1) and (2)) that we used in the proof of Theorem 4.3 express the value of an angle of a spherical triangle in terms of the three sides. There is another "cosine law" in spherical geometry which expresses a side in terms of the three angles. The two formulae are the same up to changing capital and small letters. This is due to polarity theory in spherical geometry. More precisely, to any spherical triangle ABC, there is an associated triangle $A'B'C'$ whose angles satisfy $A' = \pi - BC$, $B' = \pi - AC$ and $C' = \pi - AB$ and whose sides satisfy $A = \pi - B'C'$, $B = \pi - A'C'$ and $C = \pi - A'B'$, cf. e.g. [25] p. 3. One then uses the fact that the function cos is even.

3.— Motivated by the results expressed in Corollary 4.5, Busemann defined a notion of bounded and zero curvature in a metric space satisfying very mild properties (essentially, the fact that one can define midpoints); cf. [5], [6], and [8]. More precisely, a metric space (M, d) said to be *Busemann positively curved* if for any three points A, B, and C in M if D is a midpoint of A and B and E a midpoint of B and C, we have

$$d(D, E) > \frac{1}{2}d(A, C). \tag{5}$$

There are variations of this definition, namely, the space is called *Busemann non-negatively curved, negatively curved, or non-positively curved* depending on whether the strict inequality in (5) is replaced by a large inequality, or the reverse strict or large inequality respectively. If the inequality is replaced by an equality, the space is said to be *zero curved.* All these notions were extensively used in metric geometry after Busemann.

4.— Theorems 4.1, 4.2, and Heron's formula in Euclidean geometry show that in the three classical geometries the areas of triangles can be expressed using side lengths. A result of Busemann shows that these are the only geometries where this is possible. Indeed, he proves in [4] that a two-dimensional metric space with very mild properties is either locally isometric to the Euclidean, spherical and hyperbolic plane if and only if, locally, an area functions exists for which triangles with equal sides have equal area.

References

[1] R. Caddeo and A. Papadopoulos (ed.), *Spherical geometry in the eighteenth century: Texts and commentaries,* to appear in 2019.
☞ 8

[2] J. Beery and J. Stedall (eds.), *Thomas Harriot's doctrine of triangular numbers: The "Magisteria Magna,"* Heritage of European Mathematics, European Mathematical Society (EMS), Zürich, 2009.
MR 2516550 Zbl 1168.01001 ☞ 7

[3] R. Bonola, *La geometria non-euclidea,* Esposizione storico-critica del suo sviluppo, 1st ed., Ditta Nicola Zanchinelli editore, Bologna, 1906. German translation by M. Liebmann in the collection *Wissenschaft und Hypothese,* Teubner, Leipzig, 1908. English translation by H. S. Carslaw, *Non-Euclidean geometry,* A critical and historical study of its development. 1st ed., Chicago, 1912, reprinted by Dover, NewYork, 1955.
JFM 37.0494.02 ☞ 23

[4] H. Busemann, "Two-dimensional geometries with elementary areas," *Bull. Am. Math. Soc.* 53 (1947), 402–407. Reprinted in H. Busemann, *Selected works* (A. Papadopoulos, ed.), Vol. I, Springer Verlag, Cham, 2018, 379–384.
MR 0020257 Zbl 0038.35401 ☞ 21

[5] H. Busemann, “Spaces with non–positive curvature.” *Acta Math.* 80 (1948), 259–310. Reprinted in H. Busemann, *Selected works* (A. Papadopoulos, ed.), Vol. I, Springer Verlag, Cham, 2018, 385–436.
MR 0029531 Zbl 0038.10005 ☞ 20

[6] H. Busemann, *The geometry of geodesics,* Pure and applied Mathematics 6, Academic Press, New York, 1955.
MR 0075623 Zbl 0112.37002 ☞ 20

[7] H. Busemann and C. M. Petty, “Problems on convex bodies,” *Math. Scand.* 4 (1956), 88–94. Reprinted in H. Busemann, *Selected works* (A. Papadopoulos, ed.), Vol. I, Springer Verlag, Cham, 2018, 377–384.
MR 0084791 Zbl 0070.39301 ☞ 9

[8] H. Busemann, “The synthetic approach to Finsler spaces in the large,” C.I.M.E., Geometria del Calcolo delle Variazioni 1 (№ 2), 1961, 70 p. Reprinted in H. Busemann, *Selected works* (A. Papadopoulos, ed.), Vol. I, Springer Verlag, Cham, 2018, 597–666.
MR 2866878 Zbl 0122.40401 ☞ 20

[9] H. Busemann, *Selected works* (A. Papadopoulos, ed.), two volumes, Springer Verlag, Cham, 2018.
Zbl 06764305 (Vol. I) Zbl 06764306 (Vol. II) ☞ 3

[10] B. Cavalieri, *Directorium generale uranometricum,* in quo trigonometriæ logarithmicæ fundamenta ac regulæ demonstratur, astronomicæque supputationes ad solam fere vulgarem additionem reducentur, opus utilisissimum astronomis, mechanicis, geographicis, architectis principue militaribus, mechanicis, geographicis, nec non ipsis philosophis naturalibus; authore Fr. Bonav. Cavalioro, Mediolanensi ordinis Jesuatorum S. Hieronymi, priore titulari ac in almo Bononiensi gymnasio primario mathematicorum professore. Bonaniæ, 1632.
☞ 8

[11] L. Euler, “Principes de la trigonométrie sphérique tirés de la méthode des plus grands et des plus petits,” *Mémoires de l’Academie des Sciences de Berlin* 9 (1753), 1755, 233–257. Reprinted in *Opera omnia,* Series 1, Vol. XXVII, 277–308.
`http://eulerarchive.maa.org/pages/E214.html`
☞ 8

[12] L. Euler, “De mensura angulorum solidorum,” *Acta academiae scientarum imperialis Petropolitinae* 2, 1781, 31–54. Reprinted in *Opera omnia,* Series 1, Vol. XXVI, p. 204–223.
`http://eulerarchive.maa.org/pages/E514.html`
☞ 8

[13] L. Euler, “Variae speculationes super area triangulorum sphaericorum,” *Nova Acta Academiae Scientarum Imperialis Petropolitinae* 10 (1797), 47–62. Reprinted in *Opera omnia,* Series 1, Vol. 29, 253–266.
`http://eulerarchive.maa.org/pages/E698.html`
☞ 8, 18

[14] E. Frenkel and W. Su, “The area formula for hyperbolic triangles,” in *Eighteen essays in non-Euclidean geometry* (V. Alberge and A. Papadopoulos, eds.), Chapter 2, 27–46.
☞ 18

[15] C. F. Gauss, *Werke,* Königliche Gesellschaft der Wissenschaften, Göttingen, 1900.
☞ 13

[16] A. Girard, *Invention nouvelle en l’algèbre,* Guillaume Iansson Blaeuw, Amsterdam, 1629.
☞ 7

[17] J.-L. Lagrange, “Solution de quelques problèmes relatifs aux triangles sphériques avec une analyse complète de ces triangles,” *Journal de l’école Polytechnique* 2 (1800), 270–297.
☞ 7

[18] N. I. Lobachevsky, *Geometrische Untersuchungen zur Theorie der Parallellinien,* Kaiserl. russ. wirkl. Staatsrathe und ord, Prof. der Mathematik bei der Universität Kasan, in der Finckeschen Buchhandlung, Berlin, 1840. English translation by G. B. Halsted, *Geometrical researches on the theory of parallels,* first published by the University of Texas at Austin, 1891. Reprinted in Bonola [3].
☞ 14

[19] N. I. Lobachevsky, “Применение воображаемой геометрии к некоторым интегралам” (Application of imaginary geometry to certain integrals), *Uchenye Zapiski Kazanskogo Imperatorskogo Universiteta* 1 (1836), 3–166. Reprinted in Lobachevsky’s *Collected geometric works,* edited by the Imperial University of Kazan, Vol. I, 1886, 121–218, and in Lobachevsky’s *Complete collection of works* (V. F. Kagan, A. P. Kotelnikov, V. V. Stepanov, N. G. Chebotarev, and P. A. Shirokov, eds.), GITTL, Moscow and Leningrad, Vol. III, 1951, 181–294.
☞ 17

[20] N. I. Lobachevsky, “Géométrie imaginaire,” *J. Reine Angew. Math.* 17 (1837), 295–320.
☞ 17

[21] N. I. Lobachevsky, "Pangéométrie ou précis de géométrie fondée sur une théorie générale et rigoureuse des parallèles," in *Collection of memoirs written by professors of the University of Kazan on the occasion of the 50th anniversary of its foundation,* Vol. I, Kazan, 1856, 277–340. English edition, *Pangeometry,* translation, notes and commentary by A. Papadopoulos, Heritage of European Mathematics 4, European Mathematical Society (EMS), Zürich, 2010.
MR 2743638 Zbl 1208.51013 ☞ 17

[22] A. Papadopoulos and G. Théret, *La théorie des lignes parallèles de Johann Heinrich Lambert,* critical edition of Lambert's *Theorie der Parallellinien,* with a French translation and mathematical and historical commentaries, ed. Blanchard, coll. Sciences dans l'Histoire, Paris, 2014.
MR 3241580 Zbl 1304.01004 ☞ 13

[23] A. Papadopoulos and G. Théret, "Hyperbolic geometry in the work of J. H. Lambert," *Gaṇita Bhārātī* 36 (2014), no. 2, 129–155.
MR 3410718 Zbl 1341.01017 ☞ 13, 24

[24] A. Papadopoulos, "Three theorems of Menelaus," to appear in *Amer. Math. Monthly.*
☞ 5

[25] G. Papelier, *Éléments de trigonométrie sphérique,* Vuibert, Paris, 1930.
JFM 56.1151.03 ☞ 18, 20

[26] E. Prouhet, "Note historique sur l'aire du triangle sphérique," Bulletin de bibliographie, d'histoire et de biographie mathématiques, *Nouvelles annales de mathématiques* 14 (1855), 152–154.
☞ 8

[27] R. Raghunatan, "Review of 'La Théorie des Lignes Parallèles de Johann Heinrich Lambert' " (by A. Papadopoulos and G. Théret, see [23]), *Gaṇita Bhārātī* 36 (2014), no. 2, 205–218.
☞ 13

[28] R. Rashed and A. Papadopoulos, *Menelaus' spherics*: *early translation and al-Māhānī/al-Harawī's version,* Critical edition with English translation and historical and mathematical commentaries, Series Scientia Graeco-Arabica 21, de Gruyter, Berlin, 2017.
Zbl 06908518 ☞ 4, 5, 20

[29] P. Stäckel and F. Engel, “Gauss, die beiden Bolyai und die nichteuklidische Geometrie,” *Math. Annalen* 2 (1897), 149–167.
☞ 17

[30] P. Tannery, “Albert Girard, de Saint-Mihiel,” *Bulletin des sciences mathématiques et astronomiques,* 2^{e} série, 7 (1883), no. 1, 358–360.
☞ 7

[31] W. P. Thurston, *Three-dimensional geometry and topology* Vol. 1, edited by S. Levy, Princeton University Press, Princeton, N.J., 1997.
MR 1435975 Zbl 0873.57001 ☞ 9

[32] J. Wallis, *Operum mathematicorum,* Typis Leon, Oxford, 1693.
☞ 7

Chapter 2

The area formula for hyperbolic triangles

Elena Frenkel and Weixu Su

Contents

1. Introduction 27
2. The length element in Lobachevsky's work 28
3. Trigonometry in right triangles 31
4. An area formula by a method of Euler 36
5. The area formula revisited and Lexell's problem 43

References 45

1 Introduction

The aim of this essay is to prove some analogues in hyperbolic geometry of some of Euler's results in spherical geometry. We base our considerations on two memoirs of Euler. In the first memoir [2], Euler gives the proofs of a complete set of spherical trigonometric formulae for right triangles using a variational method. The advantage of his method is that it does not use the ambient Euclidean 3-space. Therefore, the method is in some sense intrinsic and may be used with some modifications in the setting of hyperbolic geometry.

In order to obtain the trigonometric formulae, Euler works in the so-called equidistant coordinates on the sphere and derives the length element ds in these coordinates. He uses at many places the fact that spherical geometry is infinitesimally Euclidean, i.e. that the Euclidean relations are satisfied at the level of differentials.

We work (Theorem 1.1) in the setting of Lobachevsky's geometry as introduced in [4]. There are some intersections between Lobachevsky's results and Euler's results. Both work in the same coordinates. One difference between the two approaches is that Lobachevsky uses trigonometry, which he develops from first principles, in order to find the length element ds, whereas Euler uses ds and the property of spherical geometry of being infinitesimally Euclidean, in order to reconstruct the trigonometry.

The second memoir of Euler, [3], deals with an area formula in spherical geometry in terms of side lengths and a related problem, known as *Lexell's problem*.

Let us recall Euler's area formula and the formulation of the Lexell problem in the spherical case.

Theorem 1.1 (Euler area formula). *In spherical geometry, the area $\mathcal{A}$ of a triangle with side lengths a, b, c is given by*

$$\cos\left(\frac{\mathcal{A}}{2}\right) = \frac{1+\cos(a)+\cos(b)+\cos(c)}{4\cos\left(\frac{1}{2}a\right)\cos\left(\frac{1}{2}b\right)\cos\left(\frac{1}{2}c\right)}.$$

Lexell problem. *In spherical geometry, given two distinct points A, B, determine the locus of points P such that the area of the triangle with vertices A, B, P is equal to some given constant $\mathcal{A}$.*

After proving the hyperbolic analogue of Euler's area formula (Theorem 3.1), we revisit it again and give a solution of the Lexell problem in the hyperbolic case using geometric considerations. Our proof gives a new form and a geometric interpretation to the Euler area formula.

The results and methods introduced in this chapter should be interesting if generalized to the 3-dimensional case, whether spherical or hyperbolic.

We assume in our proofs that all the functions we are dealing with are enough differentiable.

2 The length element in Lobachevsky's work

Euler introduced in [2] the coordinates on the sphere, such that $y = const$-curves are a chosen great circle G and the small circles equidistant to it, $x = const$-curves are the great circles, which are orthogonal to G. Thereafter, he deduced the formula for the length element ds in these coordinates

$$ds^2 = \cos^2(y)\,dx^2 + dy^2. \tag{1}$$

Lobachevsky proved in [4] the analogue of this formula

$$ds^2 = \cosh^2(y)\,dx^2 + dy^2. \tag{2}$$

In this section we will explain this result and introduce the related concepts and notions, which will be also of use in the proof of Theorem 1.1.

2.1 The theory of parallels and the angle of parallelism

In the *Pangeometry* Lobachevsky presents a summary of his lifelong work on hyperbolic geometry. He starts with first principles and develops from there the theory of parallel lines. He uses first the characteristic property of hyperbolic geometry: the sum of the angles in a right triangle is strictly less than two right angles. Based on this fact he defines a function Π, called *the angle of parallelism* as follows.

First, we recall the notion of parallel lines in the sense of Lobachevsky. Consider a line l and a point P not belonging to l. There is the unique perpendicular from P to the line l. We denote it by PP' and its length by p. Let $\mathcal{L}$ be the set of lines passing through P. Each line from $\mathcal{L}$ can be parametrized by the angle $\theta \in [0, \pi[$, which this line makes with the perpendicular PP' on a given side of a line containing PP'. Then *the parallel line to l through P (on a given side of PP')* is the line $l' = l_{\theta^*}$ characterized by the property that for $\theta < \theta^*$ the lines l_θ intersect l and for $\theta \geq \theta^*$, the lines l_θ and l do not intersect. Therefore, a parallel line l' to l is the limiting position of lines through P, which intersect l on a given side of P'. The angle θ^* is called *the angle of parallelism* (see Figure 1).

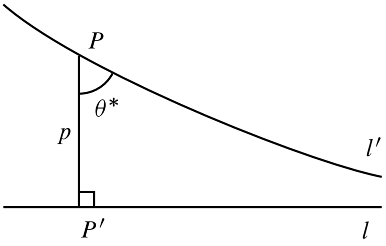


Figure 1. Angle of parallelism $\theta^* = \Pi(p)$

Definition 2.1. Let l be a line and P a point not on l. Let P' be the foot of the perpendicular dropped from P to l and p the length of PP'. A parallel line to l through P makes with the perpendicular PP' two angles, one of which is acute. This angle is called *the angle of parallelism*, and it is denoted by $\Pi(p)$.

This notation, used by Lobachevsky, emphasizes the dependence of the angle of parallelism on the length p of the perpendicular PP', which can be seen as *the angle of parallelism function*.

The first analytic formula in [4] clarifies the relation $\Pi(p)$ (see [4], p. 25):

$$\cos(\Pi(p)) = \tanh(p). \tag{3}$$

2.2 Equidistant coordinates

We introduce coordinates in the hyperbolic plane as follows. Let O be a point and OX and OY two orthogonal lines, which intersect at O. We choose a side on each line with respect to O, which we call *positive*. These lines will become x- and y-axes of the coordinate system (see Figure 2). Let M be a point in the hyperbolic plane. We denote by P_x the foot of the perpendicular from M to OX. In analogy with the usual Cartesian coordinates in the Euclidean plane, we associate to M the pair of real numbers (x, y), given by

$$x = \pm \operatorname{Length}(OP_x), \tag{4}$$

$$y = \pm \operatorname{Length}(P_x M). \tag{5}$$

The signs of x and y are chosen as follows. We take x (or y) positive, if M lies in the same halfplane bounded by the line OY (or OX) as the positive side of OX (or the positive side of OY). Likewise, we take x (or y) negative, if M and the positive side of OX(or the positive side of OY) lie in different halfplanes with boundary line OY (or OX).

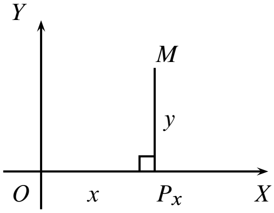


Figure 2. Equidistant coordinates (x, y)

Remark 2.2. In these coordinates, the (y = const)-lines are the line OX and the equidistant curves to OX, and the (x = const)-lines are the lines which are orthogonal to OX. Interesting facts on these coordinates systems in Euclidean, spherical and hyperbolic geometries can be found in [7].

2.3 The length element ds in coordinates

We introduce ds in equidistant coordinates (x, y), a formula due to Lobachevsky (see [4], p. 43)

$$ds = \sqrt{\frac{dx^2}{\sin^2(\Pi(y))} + dy^2}. \tag{6}$$

From (3),

$$\frac{1}{\sin(\Pi(y))} = \cosh(y).$$

Therefore, we can rewrite (6) as (2).

3 Trigonometry in right triangles

In this section we will see how to reconstruct trigonometry from ds as in (2) and from the assumption that hyperbolic geometry is at the infinitesimal level Euclidean. Practically, we assume in the proof of the following result that the hyperbolic relations limit to the Euclidean ones.

Theorem 3.1. *In the hyperbolic plane, let ABC be a right triangle with the sides a, b, c and angles α, β, $\frac{\pi}{2}$ as indicated in Figure 3. Then the following trigonometric identities hold*:

$$\sinh(b) = \frac{\tanh(a)}{\tan(\alpha)}, \quad \cosh(b) = \frac{u}{\sin(\alpha)\cosh(a)}, \quad \tanh(b) = \frac{\cos(\alpha)\sinh(a)}{u},$$
$$\sinh(c) = \frac{\sinh(a)}{\sin(\alpha)}, \quad \cosh(c) = \frac{u}{\sin(\alpha)}, \quad \tanh(c) = \frac{\sinh(a)}{u},$$
$$\sin(\beta) = \frac{\cos(\alpha)}{\cosh(a)}, \quad \cos(\beta) = \frac{u}{\cosh(a)}, \quad \tan(\beta) = \frac{\cos(\alpha)}{u},$$

where $u = \sqrt{\cosh^2(a) - \cos^2(\alpha)}$.

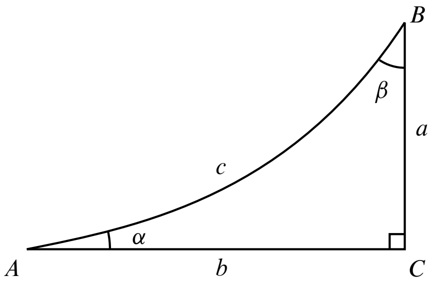


Figure 3. Right hyperbolic triangle

Proof. We work in the equidistant coordinates (x, y). We take A as the origin O of the coordinate system and the line containing AC as OX. The positive side of OX (with respect to O) is a side containing the point C. We choose the positive side of OY such that the ordinate of the point B is positive.

The line AB in coordinates (x, y) can be seen as the graph of a function

$$g\colon [0, b] \longrightarrow \mathbb{R},$$

that is,

$$AB = \text{Graph}(g) = \{(x, g(x)) \mid x \in [0, b]\}.$$

This provides a parametrization of AB by $\gamma\colon [0, b] \to \mathbb{R}^2$, given by $x \mapsto (x, g(x))$, and the length of AB can be computed by means of ds (compare with (2)):

$$\text{Length (AB)} = \int_0^b \sqrt{\cosh^2(g(x)) + (g'(x))^2}\, dx. \tag{7}$$

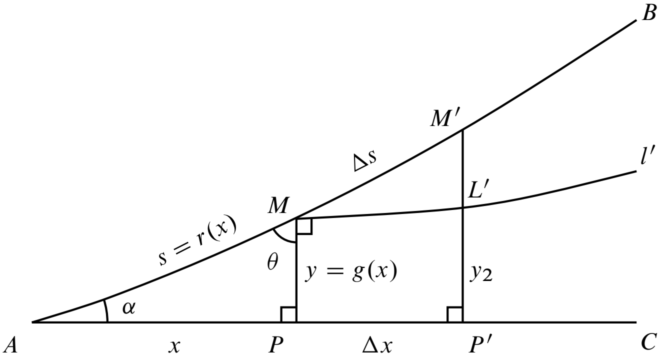


Figure 4. Proof of the theorem

Since $g(b) = a$, the function g is useful in the search for the relations between the sides of the triangle ABC. We used in (7) implicitly another function of this type, $r\colon [0, b] \to [0, c]$, given by $r(x) = \text{Length}(AM)$, where M has the coordinates $(x, g(x))$. We have $r(b) = c$. The function under the integral sign in (7) is the derivative of r,

$$r'(x) = \sqrt{\cosh^2(g(x)) + (g'(x))^2}. \tag{8}$$

The idea of the proof is to deduce a differential equation in terms of $g(x)$ and $g'(x)$ using the calculus of variations. This equation combined with (8) will allow us to find $r(b)$ and the other relations of this theorem.

Among all the curves connecting A and B, which can be represented in coordinates as graphs of differentiable functions, the line segment AB has the minimal length. Therefore, the function $g(x)$ minimizes the length functional L, given by

$$L(f) = \text{Length}(c_f) = \int_0^b \sqrt{\cosh^2(f(x)) + (f'(x))^2}\,dx \tag{9}$$

and defined on the functional space

$$C^1_{0,a}([0,b],\,\mathbb{R}) = \{f\colon [0,b] \to \mathbb{R} \mid f \text{ is differentiable, } f(0) = 0 \text{ and } f(b) = c\}.$$

In (9), we denote by c_f a curve connecting A and B, such that it can be represented in coordinates as Graph(f).

The differential of L vanishes necessarily at g and in particular we have

$$\frac{d}{d\varepsilon} L(g + \epsilon h)|_{\epsilon=0} = 0 \tag{10}$$

for every differentiable function $h\colon [0,b] \to \mathbb{R}$ with $g + \varepsilon h \in C^1_{0,a}([0,b],\,\mathbb{R})$.

For the sake of brevity, we introduce a function $F\colon \mathbb{R}^2 \to \mathbb{R}$, given by

$$F(x_1, x_2) = \sqrt{\cosh^2(x_1) + x_2^2},$$

such that the functional L becomes

$$L(f) = \int_0^b F(f(x), f'(x))\,dx.$$

The directional derivative in the left-hand side of (10) after integration by parts becomes

$$\frac{d}{d\varepsilon} L(g + \varepsilon h)|_{\varepsilon=0} = \int_0^b \Big(\frac{\partial F}{\partial x_1}(g(x), g'(x)) - \frac{d}{dx}\frac{\partial F}{\partial x_2}(g(x), g'(x))\Big) h\,dx.$$

This relation vanishes if and only if the term in brackets is identically zero. We obtain this result by taking a positive function φ for h. The next relation was named later on *Euler–Lagrange formula*

$$\frac{\partial F}{\partial x_1}(g(x), g'(x)) - \frac{d}{dx}\frac{\partial F}{\partial x_2}(g(x), g'(x)) = 0, \quad \text{for all } x \in [0,b]. \tag{11}$$

We simplify further the notation by means of the functions

$$G_i(x) = \frac{\partial F}{\partial x_i}(g(x), g'(x)), \quad i = 1, 2. \tag{12}$$

The Euler–Lagrange formula (11) gives the simple relation

$$G_1 = G_2'. \tag{13}$$

On the other hand, we manipulate the function $r'(x)$, which stands under the integral sign in $L(g)$ (compare with (8))

$$r''(x) = \frac{\partial F}{\partial x_1}(g(x), g'(x))\, g'(x) + \frac{\partial F}{\partial x_2}(g(x), g'(x))\, g''(x),$$

or in terms of the functions G_i and using further the relation for G_i, given by the Euler–Lagrange formula (13)

$$r'' = G_1\, g' + G_2\, g'' = G_2'\, g' + G_2\, g'' = (G_2\, g')',$$

we finally get

$$r' = G_2\, g' + C,$$

where C is a constant with respect to the variable x.

We rewrite this relation using the definition of G_2 (see (12)) and the relation $r'(x)$ (see (8))

$$\sqrt{(g')^2 + \cosh^2(g)} = \frac{(g')^2}{(g')^2 + \cosh^2(g)} + C.$$

Resolving this equation in g' gives

$$g'(x) = \frac{\cosh(g(x))\sqrt{\cosh^2(g(x)) - C^2}}{C} \tag{14}$$

and (8) takes the form

$$r'(x) = \frac{\cosh^2(g(x))}{C}. \tag{15}$$

Let M and M' be two points on AB with coordinates (x, y) and (x', y'), respectively. We denote by P and P' the feet of the perpendiculars dropped from M and M'. Let l' be a halfline, orthogonal to MP. We denote the intersection of l' with $M'P'$ by L'. We obtain the trirectangular quadrilateral $PML'P'$ with the acute angle L'. We denote the length of $L'P'$ by y_2, the length of AM by s, which coincides with $r(x)$, the length of MM' by Δs, $x' - x$ by Δx and $y' - y$ by Δy (see Figure 4).

Taking $M = A$, we determine the value of a constant C from a limit Euclidean relation for $\cos\alpha$ in the triangle $AM'P'$ (see Figure 5).

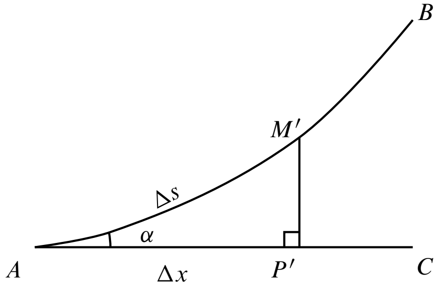


Figure 5. Case $M = A$

In a right hyperbolic triangle holds a limit Euclidean relation:

$$\cos(\alpha) = \lim_{\Delta s \to 0} \frac{\Delta x}{\Delta s} = \frac{1}{r'(0)} = C.$$

We obtain

$$C = \cos(\alpha). \tag{16}$$

We return to the general situation again, i.e. M is an arbitrary point on AB. We denote the angle $\angle AMP$ by θ. In the triangle $MM'L'$, the angle $\angle M'ML' = \frac{\pi}{2} - \theta$. If Length$(MM') \to 0$, the angle $\angle M'L'M$ tends to $\frac{\pi}{2}$ and for $MM'L'$ holds a limit Euclidean relation for right triangles

$$\sin(\angle M'ML') = \lim_{\Delta s \to 0} \frac{y' - y_2}{\Delta s}.$$

The value of $\theta = \angle AMP$ depends on the position of M. We define a function $w\colon [0, b] \to [0, \beta]$, such that $w(x) = \angle AMP$,

$$\begin{aligned} \cos(w(x)) &= \lim_{\Delta s \to 0} \frac{y' - y_2}{\Delta s} \\ &= \lim_{\Delta s \to 0} \frac{\Delta y}{\Delta s} \\ &= \frac{g'(x)}{r'(x)} \\ &= \frac{\sqrt{\cosh^2(g(x)) - C^2}}{\cosh(g(x))}. \end{aligned}$$

Taking $x = b$ in this relation and replacing C by (16), we get

$$\cos(\beta) = \frac{\sqrt{\cosh^2(a) - \cos^2(\alpha)}}{\cosh(a)}. \tag{17}$$

We get the relations for $r(b)$ and $g^{-1}(a)$ by means of (14) and (15),

$$c = r(b) = \int_0^b r'(x)\,dx = \int_0^a \frac{\cosh(g)}{\sqrt{\cosh^2(g) - C^2}}\,dg = \operatorname{acosh}\left(\frac{\cosh^2(a) - C^2}{1 - C^2}\right),$$

$$b = \int_0^b dx = \int_0^a \frac{C}{\cosh(g)\sqrt{\cosh^2(g) - C^2}}\,dg = \operatorname{asinh}\left(\frac{C\tanh(a)}{\sqrt{1 - C^2}}\right).$$

Replacing C by (16), we finally obtain

$$\cosh(c) = \frac{\sqrt{\cosh^2(a) - \cos^2(\alpha)}}{\sin(\alpha)}, \tag{18}$$

$$\sinh(b) = \frac{\cos(\alpha)\tanh(a)}{\sin(\alpha)} = \frac{\tanh(a)}{\tan(\alpha)}. \tag{19}$$

The remaining relations of Theorem 3.1 follow easily from (17)–(19) and from the properties of the functions cosh, sinh and cos. ☐

4 An area formula by a method of Euler

Theorem 4.1 (hyperbolic Euler Formula). *In the hyperbolic plane, the area $\mathcal{A}$ of a triangle with side lengths a, b, c is given by*

$$\cos\left(\frac{\mathcal{A}}{2}\right) = \frac{1 + \cosh(a) + \cosh(b) + \cosh(c)}{4\cosh\left(\frac{1}{2}a\right)\cosh\left(\frac{1}{2}b\right)\cosh\left(\frac{1}{2}c\right)}.$$

Beforehand, we deduce a formula for the area of the hyperbolic circular region $K_{r,\varphi}$ with radius r and angle φ (see Figure 6), which we will use in the proof.

For this, we cite the result of J.-M. de Tilly from his work on hyperbolic geometry [6], see also [5] in this book for the description of a kinematic approach of de Tilly. He introduced a function $\operatorname{circ}(r)$, which is defined as the circumference of the hyperbolic circle of radius r and showed without use of trigonometry that

$$\operatorname{circ}(r) = 2\pi\sinh(r). \tag{20}$$

Lemma 4.2. *The area of the hyperbolic circular region $K_{r,\varphi}$ with radius r and angle φ is*

$$\operatorname{Area}(K_{r,\varphi}) = |\varphi|(\cosh(r) - 1).$$

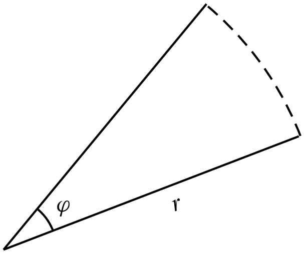


Figure 6. Region $K_{r,\varphi}$

Proof. By (20), the circumference of the hyperbolic circle of radius x is $2\pi\sinh(x)$ and hence, the length $l(x)$ of the hyperbolic circular arc of the same radius and angle φ is $\sinh x|\varphi|$. Then, for the area of circular segment, we have

$$\operatorname{Area}(K_{r,\varphi}) = |\varphi| \int_0^r \sinh(x)\,dx = |\varphi|(\cosh(r)-1). \qquad \square$$

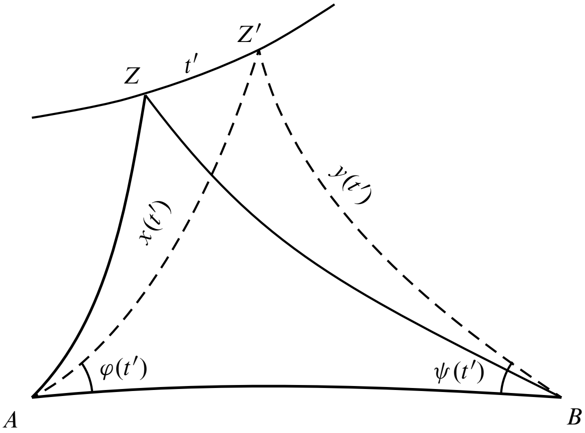


Figure 7. Proof of Theorem 1.1

Proof of Theorem 1.1. Let AB be a segment of fixed length a. Let Z be a point in a chosen connected component of $\mathbb{H}^2 \backslash \{l\}$, where l is a line containing A and B. We fix a ray r, starting at Z and consider the area function $\mathcal{A}\colon [0,\epsilon) \to \mathbb{R}_{>0}$, given by $\mathcal{A}(t') = \operatorname{Area}(ABZ')$, where Z' is a point on r such that the length of ZZ' is t'.

Besides the function $\mathcal{A}$, we define in the same manner the functions x, y, φ and ψ, such that $x(t')$, $y(t')$, $\varphi(t')$ and $\psi(t')$ are the values of the side lengths AZ', BZ' and of the angles $\angle Z'AB$ and $\angle ABZ'$, respectively (see Figure 7).

We are looking for the expression for the differential $d\mathcal{A}$. We consider the area change

$$\Delta\mathcal{A}(t') = \text{Area } (ABZ') - \text{Area } (ABZ).$$

The area change is given by

$$\Delta\mathcal{A} = \operatorname{sgn}(\Delta\varphi)\,\text{Area}(AZZ') + \operatorname{sgn}(\Delta\psi)\,\text{Area}(BZZ')\,. \tag{21}$$

In order to see this, we distinguish the four cases:

1. $\operatorname{sgn}(\Delta\varphi) = \operatorname{sgn}(\Delta\psi) = 1$,
2. $\operatorname{sgn}(\Delta\varphi) = \operatorname{sgn}(\Delta\psi) = -1$,
3. $\operatorname{sgn}(\Delta\varphi) = -1$, $\operatorname{sgn}(\Delta\psi) = 1$,
4. $\operatorname{sgn}(\Delta\varphi) = 1$, $\operatorname{sgn}(\Delta\psi) = -1$.

Since $\Delta\mathcal{A}$ corresponding to the cases 1 and 3 is equal to $-\Delta\mathcal{A}$ in the cases 2 and 4, respectively, we reduce our consideration to the cases 1 and 3.

In the case 1, one triangle is entirely contained in the other one and formula (21) can be easily seen from Figure 8.

In the case 3 a side of one triangle intersects a side of another triangle. We denote the intersection point by O (see Figure 7 for the case 3).

First, the area difference does not change, if one removes from the triangles ABZ' and ABZ the common triangle AOB

$$\text{Area}(ABZ') - \text{Area}(ABZ) = \text{Area}(BOZ') - \text{Area}(AOZ).$$

Further, adding ZOZ' to both triangles BOZ' and AOZ does not affect the area difference again, therefore we finally get

$$\text{Area}(ABZ') - \text{Area}(ABZ) = \text{Area}(BZZ') - \text{Area}(AZZ'),$$

which is formula (21) for the case 3.

We consider

$$\lim_{t'\to+0}\frac{\Delta\mathcal{A}(t')}{t'} = \operatorname{sgn}(\Delta\varphi)\lim_{t'\to+0}\frac{\text{Area}(AZZ')}{t'} + \operatorname{sgn}(\Delta\psi)\lim_{t'\to+0}\frac{\text{Area}(BZZ')}{t'}. \tag{22}$$

Since the area of AZZ' is enclosed by the areas of the hyperbolic circular segments $K_{x(0),\varphi}$ and $K_{x(t'),\varphi}$, by Lemma 4.2 we have an estimate

$$|\Delta\varphi|(\cosh(x(0)) - 1) \le \text{Area}(AZZ') \le |\Delta\varphi|(\cosh(x(t')) - 1),$$

from which it follows that

$$\lim_{t'\to+0}\frac{\mathrm{Area}(AZZ')}{t'} = |\Delta\varphi|\,(\cosh(x(0)) - 1). \tag{23}$$

For the sake of brevity, we denote $x(0)$ by x and $y(0)$ by y.

Combining (23) with (22), we get the following relations in terms of the differentials $d\mathcal{A}$, $d\varphi$ and $d\psi$

$$d\mathcal{A} = (\cosh(x) - 1)\,d\varphi + (\cosh(y) - 1)\,d\psi. \tag{24}$$

In order to rewrite this relation in terms of the differentials of the sides dx and dy, we use the First Cosine Law. Applying it to ABZ', we obtain

$$\cos(\varphi(t')) = \frac{\cosh(a)\cosh(x(t')) - \cosh(y(t'))}{\sinh(a)\sinh(x(t'))}.$$

By differentiation, we get

$$-\sin(\varphi)\,d\varphi = \frac{(-\cosh(a) + \cosh(y)\cosh(x))\,dx - \sinh(x)\sinh(y)\,dy}{\sinh(a)\sinh^2(x)},$$

where x, y, and φ denote $x(0)$, $y(0)$, and $\varphi(0)$, respectively.

From now on and until the relation (28), we follow the computations of Euler, in order to find a well-suited substitution $v = f(a, x, y)$ for the integration of $d\mathcal{A}$.

Since

$$\sin(\varphi) = \sqrt{1 - \cos^2(\varphi)} = \frac{w}{\sinh(a)\sinh(x)},$$

where

$$w := \sqrt{1 - \cosh^2(x) - \cosh^2(a) - \cosh^2(y) + 2\cosh(a)\cosh(x)\cosh(y)},$$

we have

$$d\varphi = \frac{(\cosh(a) - \cosh(y)\cosh(x))\,dx + \sinh(x)\sinh(y)\,dy}{\sinh(x)w}.$$

Analogously, we obtain

$$d\psi = \frac{(\cosh(a) - \cosh(x)\cosh(y))\,dy + \sinh(x)\sinh(y)\,dx}{\sinh(y)w}.$$

Therefore (24) becomes

$$\begin{aligned} d\mathcal{A} = {} & \frac{(\cosh(a) - \cosh(y)\cosh(x))\,dx + \sinh(x)\sinh(y)dy}{\sinh(x)w}(\cosh(x) - 1) \\ & + \frac{(\cosh(a) - \cosh(x)\cosh(y))\,dy + \sinh(x)\sinh(y)dx}{\sinh(y)w}(\cosh(y) - 1). \end{aligned}$$

Now, using the trigonometric formulae

$$\tanh\left(\frac{x}{2}\right) = \frac{\cosh(x)-1}{\sinh(x)} = \frac{\sinh(x)}{\cosh(x)+1},$$

we simplify our equation

$$\begin{aligned} w\,d\mathcal{A} &= \tanh\left(\frac{x}{2}\right)(\cosh(a)+\cosh(y)-\cosh(x)-1)\,dx \\ &\quad + \tanh\left(\frac{y}{2}\right)(\cosh(a)+\cosh(x)-\cosh(y)-1)\,dy. \end{aligned}$$

Taking the symmetric expression

$$s := \cosh(a)+\cosh(x)+\cosh(y),$$

we get

$$\begin{aligned} w\,d\mathcal{A} &= (-1+s)\Big(\tanh\left(\frac{x}{2}\right)dx + \tanh\left(\frac{y}{2}\right)dy\Big) \\ &\quad - 2\tanh\left(\frac{x}{2}\right)\cosh(x)\,dx - 2\tanh\left(\frac{y}{2}\right)\cosh(y)\,dy. \end{aligned}$$

Applying again the (former) expression of $\tanh\left(\frac{x}{2}\right)$, we have

$$\tanh\left(\frac{x}{2}\right)\cosh(x) = -\tanh\left(\frac{x}{2}\right) + \sinh(x),$$

so that

$$\frac{w\,d\mathcal{A}}{1+s} = \tanh\left(\frac{x}{2}\right)dx + \tanh\left(\frac{y}{2}\right)dy - 2\frac{\sinh(x)\,dx+\sinh(y)\,dy}{1+s}. \tag{25}$$

Using $\int \tanh\left(\frac{1}{2}x\right)dx = 2\ln\left(\cosh\left(\frac{1}{2}x\right)\right)$ and

$$\int \frac{\sinh(x)\,dx+\sinh(y)\,dy}{1+s} = \int \frac{d(1+s)}{1+s} = \ln(1+s),$$

we can find the antiderivative of the right-hand side of the equation (25) and we get

$$\int \frac{w\,d\mathcal{A}}{1+s} = 2\ln\frac{q}{(1+s)\cosh\left(\frac{a}{2}\right)}, \tag{26}$$

where q is given by the symmetric expression

$$q := \cosh\left(\frac{a}{2}\right)\cosh\left(\frac{x}{2}\right)\cosh\left(\frac{y}{2}\right).$$

Taking $p := \frac{q}{1+s}$ and differentiating the equation, we obtain

$$\frac{w\,d\mathcal{A}}{1+s} = 2\frac{dp}{p},$$

or equivalently,

$$d\mathcal{A} = \frac{2dp}{p}\frac{1+s}{w}. \tag{27}$$

The term $\frac{1+s}{w}$ can be expressed as a function of p. Indeed, we note that

$$\begin{aligned} w^2 + (1+s)^2 &= 2(1+\cosh(a)(1+\cosh(x))(1+\cosh(y)) \\ &= 16\cosh^2\left(\frac{a}{2}\right)\cosh^2\left(\frac{x}{2}\right)\cosh^2\left(\frac{y}{2}\right) \\ &= 16q^2. \end{aligned}$$

Hence,

$$w = \sqrt{16q^2 - (1+s)^2}$$

and therefore

$$\frac{w}{1+s} = \sqrt{16p^2 - 1}.$$

Then (27) takes the form

$$d\mathcal{A} = \frac{2dp}{p\sqrt{16p^2-1}}.$$

Letting now $v = \frac{1}{4p}$, we obtain

$$d\mathcal{A} = -\frac{2dv}{\sqrt{1-v^2}}. \tag{28}$$

In terms of v, the primitive of $d\mathcal{A}$ can be computed, so that we get

$$\begin{aligned} A &= C + 2\arccos(v) \\ &= C + 2\arccos\left(\frac{1+\cosh(a)+\cosh(x)+\cosh(y)}{4\cosh\left(\frac{1}{2}a\right)\cosh\left(\frac{1}{2}x\right)\cosh\left(\frac{1}{2}y\right)}\right). \end{aligned}$$

We determine the constant C by considering the degenerate Euclidean case $\mathcal{A} = 0$. If $x = a$, we get $y = 0$, such that the form of our equation becomes

$$0 = C + 2\arccos\left(\frac{2+2\cosh(a)}{4\cosh^2\left(\frac{1}{2}a\right)}\right).$$

From $4\cosh^2\left(\frac{1}{2}a\right) = 2 + 2\cosh(a)$ follows $C = 0$.

Taking now $x = b$ and $y = c$, we finally get the result

$$\cos\left(\frac{1}{2}\mathcal{A}\right) = \frac{1 + \cosh(a) + \cosh(b) + \cosh(c)}{4\cosh\left(\frac{a}{2}\right)\cosh\left(\frac{b}{2}\right)\cosh\left(\frac{c}{2}\right)}. \qquad \square$$

Lemma 4.3. *The solution of the Lexell problem is a smooth curve.*

Proof. Let $\mathcal{P}_1$ be a chosen connected component of $\mathbb{H}\backslash\{l\}$, where l contains the points A and B. The solution of the Lexell problem is the set

$$L = \{Z \in \mathcal{P}_1 \mid \text{Area}(ABZ) = \mathcal{A}\}.$$

The set L is a smooth curve if the area function $\mathcal{B}\colon \mathcal{P}_1 \to \mathbb{R}_{>0}$, given by

$$\mathcal{B}(Z) = \text{Area}(ABZ)$$

is regular or, equivalently, for every point $Z \in \mathcal{P}_1$ there is a ray r starting at Z, such that the value of $\mathcal{B}$ changes in the direction of this ray.

For a fixed $Z \in \mathcal{P}_1$, consider $Z' \in \mathcal{P}_1$ in the interior of $\angle A'ZB'$, where the points A' and B' belong to the extensions of the sides AZ and BZ respectively (see Figure 8). Using Pasch's axiom of neutral geometry, one can deduce that ABZ is contained in ABZ', which gives $\mathcal{B}(Z') > \mathcal{B}(Z)$. This gives that the area value changes along ZZ'. Therefore, the function $\mathcal{B}$ is regular. $\square$

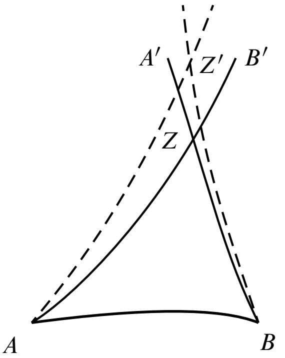


Figure 8. Area function $\mathcal{A}$ is regular

Lemma 4.3 also follows from the fact that the area of triangles is a function that depends real analytically on the three vertices and that it is non-degenerate.

5 The area formula revisited and Lexell's problem

We start this section by recalling the Girard theorem, which states that the area of a spherical triangle on a sphere of radius 1 is the angular excess, i.e. the (positive) difference between the sum of the angles of this triangle and π. In the case of the hyperbolic plane, the following is the hyperbolic analogon of this theorem.

Theorem 5.1 (hyperbolic Girard theorem). *In the hyperbolic plane, the area of a triangle with angles* α, β *and* γ *is given by the angular deficit*

$$\mathcal{A} = \pi - (\alpha + \beta + \gamma).$$

A proof of this theorem without use of trigonometry can be found in [1], Section 5.3.

In his proof of the area formula, Euler fixes the length of one side of triangle, being inspired by the Lexell problem, where one is looking for the configuration set of all triangles with the same base and the same area. The quantity v, which appears at the end of the proof of Theorem 3.1 has the property that it varies or remains constant at the same time with the area $\mathcal{A}$, according to the form of (28). Recall that

$$v = \frac{1 + \cosh(a) + \cosh(b) + \cosh(c)}{4\cosh\left(\frac{a}{2}\right)\cosh\left(\frac{b}{2}\right)\cosh\left(\frac{c}{2}\right)}. \tag{29}$$

In order to give a geometric interpretation of the Euler formula, one is led then to look for quantities which will not be affected by the change of the area.

The crucial construction here is the dissection of the hyperbolic triangle explained in [1], Chapter 5.5.

Before giving this construction, let us note that we use here the term *congruence* of triangles or more generally, of plane figures F_1 and F_2, as a primary notion of hyperbolic geometry, developed from the first principles (see [1] Ch. 2 for more details). We will write $F_1 \equiv F_2$ for this relation. Moreover, we say that two figures F_1 and F_2 are *scissors equivalent*, if one of these figures can be cut into a finite number of pieces and reassembled into the other one. It is easy to show that the relation of scissors congruence defines an equivalence relation on plane figures and that the scissors equivalent figures have the same area, in whatever reasonable sense area is defined. (In fact, in Euclid's *Elements* the notion of "having the same area" is defined as scissors equivalence of figures.)

Let ABZ be a triangle in the hyperbolic plane, A' and B' be the midpoints of AZ and BZ respectively, and F, G, H be the feet of the perpendiculars dropped from A, Z, B respectively on the line $A'B'$ (see Figure 9). By construction, the following congruences hold: $AFA' \equiv A'GZ$ and $ZGB' \equiv B'BH$. From this one deduces easily that the triangle ABZ is scissors equivalent to the *Khayyam-Saccheri quadrilateral* $BHFA$ (i.e., a quadrilateral, which has two opposite edges of the same length and makes right angles with the third common edge ([1], Definition 3.28)).

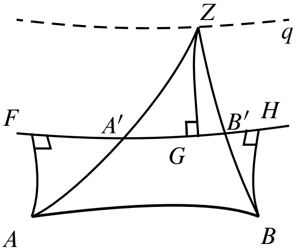


Figure 9. Dissection of hyperbolic triangle and the Lexell problem

This construction is closely related to the solution of the Lexell problem. Indeed, consider a triangle with notation as in Figure 9. Let $\mathcal{A}$ denote the area of this triangle. Let q be an equidistant curve to the line $A'B'$, which contains Z. Then every triangle ABZ' such that Z' lies on q and close to Z has the same area $\mathcal{A}$. This follows immediately from the consideration that all such triangles are scissors equivalent to the same quadrilateral $BHFA$. By Lemma 4.3, the solution of the Lexell problem is a smooth curve. Thus, the solution is exactly the curve q.

Let m denote the hyperbolic length of $A'B'$. By construction, in the quadrilateral $BHFA$ the length of HF is $2m$, and the length of AF is the same as the length of BH and is equal to h and the length of AB is a. Therefore, these quantities can be related to v, since they remain constant if the area stays constant. A direct computation shows that

$$\cosh(m) = \frac{1+\cosh(a)+\cosh(x)+\cosh(y)}{4\cosh\left(\frac{x}{2}\right)\cosh\left(\frac{y}{2}\right)}.$$

Thus the quantity v turns out to be

$$v = \frac{\cosh(m)}{\cosh\left(\frac{a}{2}\right)}$$

and we get immediately an alternative elegant expression for the hyperbolic Euler formula

$$\cos\left(\frac{\mathcal{A}}{2}\right) = \frac{\cosh(m)}{\cosh\left(\frac{a}{2}\right)}. \tag{30}$$

As a conclusion, we will give a geometric interpretation of (30) on the quadrilateral $BHFA$. The perpendicular bisector of AB divides the Khayyam–Saccheri quadrilateral $BHFA$ into two congruent trirectangular quadrilaterals ([1], Chapter 3.6).

On the one hand, by the dissection argument, the area of such a quadrilateral is equal to $\frac{\mathcal{A}}{2}$ and on the other hand, by the Hyperbolic Girard theorem, it is given by the angle defect $2\pi - \left(\alpha + 3\frac{\pi}{2}\right) = \frac{\pi}{2} - \alpha$. So we have

$$\cos\left(\frac{\mathcal{A}}{2}\right) = \cos\left(\frac{\pi}{2} - \alpha\right) = \sin(\alpha).$$

Thus the formula (30) reflects the trigonometrical relation in the trirectangular quadrilaterals

$$\sin(\alpha) = \frac{\cosh(m)}{\cosh\left(\frac{a}{2}\right)},$$

which can be found in [1], Chapter 6.6. The proof is a trigonometric computation which uses the trigonometric formulas on triangles introduced in this chapter.

References

[1] N. A'Campo and A. Papadopoulos, "Notes on non-Euclidean geometry," in *Strasbourg master class on geometry* (A. Papadopoulos, ed.), IRMA Lectures in Mathematics and Theoretical Physics 18, European Mathematical Society (EMS), Zürich, 2012, 1–182.
MR 2931886 Zbl 1253.51001 ☞ 43, 44, 45

[2] L. Euler, "Principes de la trigonométrie sphérique tirés de la méthode des plus grands et des plus petits," *Mémoires de l'Académie des sciences de Berlin* 9 (1753), 1755, 233–257. Reprinted in *Opera omnia,* Series 1, Vol. XXVII, 277–308.
`http://eulerarchive.maa.org/pages/E214.html`
☞ 27, 28

[3] L. Euler, "Variae speculationes super area triangulorum sphaericorum," *Nova Acta academiae scientarum imperialis Petropolitinae* 10, 1797, 47–62. Reprint in *Opera omnia,* Series 1, Vol. XXIX, 253–266.
`http://eulerarchive.maa.org/pages/E698.html`
☞ 28

[4] N. I. Lobachevsky, "Pangéométrie ou précis de géométrie fondée sur une théorie générale et rigoureuse des parallèles," in *Collection of memoirs written by professors of the University of Kazan on the occasion of the 50th anniversary of its foundation,* Vol. I, Kazan, 1856, 277–340. English edition, *Pangeometry,* translation, notes and commentary by A. Papadopoulos, Heritage of European Mathematics 4, European Mathematical Society (EMS), Zürich, 2010.
MR 2743638 Zbl 1208.51013 ☞ 27, 28, 29, 30

[5] D. Slutskiy, "De Tilly's mechanical view on hyperbolic and spherical geometries," in *Eighteen essays in non-Euclidean geometry* (V. Alberge and A. Papadopoulos, eds.), European Mathematical Society (EMS), Zürich, 2019, Chapter 7, 93–111.
☞ 36

[6] J.-M. de Tilly, *Études de mécanique abstraite.* Mémoires couronnée et autres textes autres mémoires Académie Royale de Belgique XXI, 1870.
☞ 36

[7] E. B. Vinberg, "Объемы неевклидовых многогранников" ("Volumes of non-Euclidean polyhedra"), *Uspekhi Mat. Nauk* 48 (1993), no. 2(290), 17–46, English translation in *Russian Math. Surveys* 48 (1993), no. 2, 15–45.
MR 1239859 Zbl 0799.01019 ☞ 30

Chapter 3

On a problem of Schubert in hyperbolic geometry

Vincent Alberge and Elena Frenkel

Contents

1. Introduction 47
2. Background in hyperbolic geometry 49
3. An elementary solution 51
4. The use of the Lexell curve in hyperbolic geometry 53

References 55

1 Introduction

This essay is motivated by the paper "Problematis cuiusdam sphaerici solutio" (Solution of a problem on the sphere) [8] of F. T. von Schubert,[1] where he solves an extremal problem in spherical geometry, namely, for a triangle of given base and fixed altitude, he proves that the minimal area is attained at the point whose the foot of the corresponding altitude is the midpoint of the base, and the maximal area is attained at the point which gives the minimal area for the antipodal given base. See Figure 1 for more details.

The purpose of this essay is to solve the equivalent problem in hyperbolic geometry. More precisely, let A and B be two fixed points on a geodesic line $\mathcal{G}$ in the hyperbolic plane, and $\mathcal{E}(h)$ be a connected component of the set of points which are at distance h from $\mathcal{G}$. Such a component is called a *hypercycle* and the geodesic line $\mathcal{G}$ is called the *axis* of $\mathcal{E}(h)$. Thus, any point which lies on this hypercycle defines a triangle whose base is AB and altitude is h. Therefore, we are interested in finding the point on $\mathcal{E}(h)$ which determines the triangle with the maximal or minimal area. The hyperbolic analogue of Schubert's result is:

---

[1] Friedrich Theodor von Schubert (1758–1825) was a German geographer and geometer, and he was a successor of Euler at the Saint Petersburg Academy of Sciences.

Theorem 1. *Among the hyperbolic triangles with base* AB *and third vertex belonging to the hypercycle* $\mathcal{E}(h)$*, the isosceles triangle is the unique triangle with maximal area, and it is the only one with extremal area.*

In order to give a proof, we organized this essay as follows. We shall start by recalling some basic notions in hyperbolic geometry such as trigonometric formulae, area formulae and the Poincaré upper half-plane model. We shall then prove in an elementary way that there is only one critical point for the corresponding area function, a critical point which solves our problem. Finally, we shall give another proof by using a curve which is the analogue in hyperbolic geometry of the so-called Lexell curve. Such a curve in the hyperbolic setting, was defined in [1] and it is also studied in the present volume by the second author in a chapter written in collaboration with Weixu Su (see [4]). Roughly speaking, the Lexell curve represents the level set of the area function of triangles where the base is fixed.

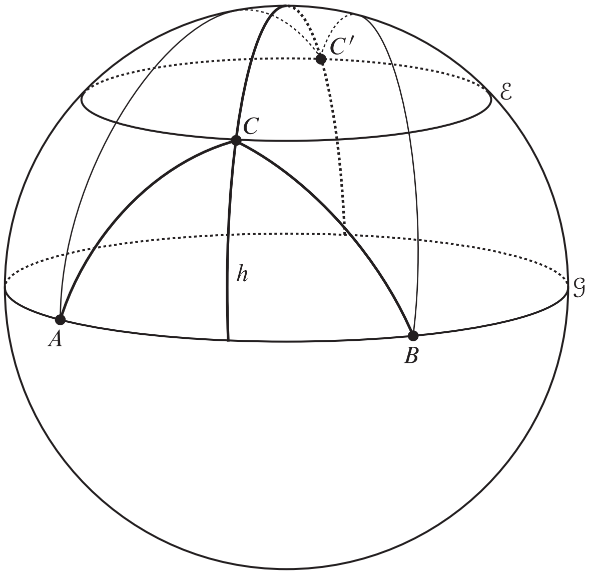


Figure 1. A geodesic line $\mathcal{G}$ with two points A and B and an equidistant curve $\mathcal{E}\ (h)$ at (spherical) distance h from $\mathcal{G}$. The points C and C' such that $AC = CB$ and $AC' = BC'$ give the triangles ABC and ABC' with minimal and maximal area, respectively. Furthermore, C and C' are antipodal on the small circle $\mathcal{E}\ (h)$.

2 Background in hyperbolic geometry

This section can be skipped by the reader familiar with hyperbolic geometry. We shall only give necessary formulae and backgrounds for the proof of Theorem 1.

We recall that the area of a hyperbolic triangle ABC is given by the angular deficit

$$\mathrm{Area}(ABC) = \pi - \widehat{BAC} - \widehat{ABC} - \widehat{ACB}. \tag{1}$$

For an elementary proof we refer to [2] in this present volume. Hence, if the triangle is right at A we have

$$\mathrm{Area}(ABC) = \frac{\pi}{2} - \widehat{ABC} - \widehat{ACB}. \tag{2}$$

We continue with hyperbolic right-angled triangles by giving some trigonometric relations. Let ABC be a triangle which is right at A and whose lengths AB and AC are c and h, respectively. If we set $\gamma = \widehat{ACB}$ and $\beta = \widehat{ABC}$ (see Figure 2), then

$$\tan(\gamma) = \frac{\tanh(c)}{\sinh(h)} \quad \text{and} \quad \tan(\beta) = \frac{\tanh(h)}{\sinh(c)}. \tag{3}$$

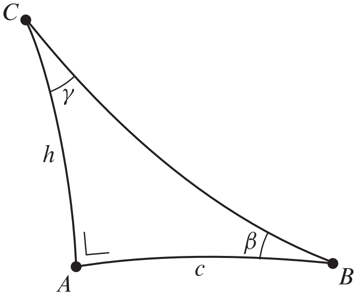


Figure 2. A right-angled hyperbolic triangle

It is worth noting that relation (3) can be proven in many different ways. In particular, a proof which uses a kinematic approach, can be found in [3] by J. M. de Tilly. See also the essay by Slutskiy [9] in the present volume.

A straightforward application of relations (2) and (3) is the following.

Lemma 2 (area formula). *Let T be a right-angled hyperbolic triangle whose base length is c and altitude h. Then*

$$\mathrm{Area}(T) = 2\arctan\left(\tanh\left(\frac{h}{2}\right)\tanh\left(\frac{c}{2}\right)\right).$$

Although this is elementary, for the convenience of the reader we provide a proof.

Proof. Let ABC be a right-angled triangle as in Figure 2. From classical relations between hyperbolic[2] and trigonometric functions and from relations (2) and (3) we get

$$\begin{aligned}
\tan\left(\mathrm{Area}(ABC)\right) &= \frac{1}{\tan\left(\gamma+\beta\right)} \\
&= \frac{1-\tan\left(\gamma\right)\tan\left(\beta\right)}{\tan\left(\gamma\right)+\tan\left(\beta\right)} \\
&= \frac{1-\frac{1}{\cosh(h)\cosh(c)}}{\frac{\tanh(h)}{\sinh(c)}+\frac{\tanh(c)}{\sinh(h)}} \\
&= \frac{\sinh\left(h\right)\sinh\left(c\right)\left(\cosh\left(h\right)\cosh\left(c\right)-1\right)}{\sinh^2\left(h\right)\cosh\left(c\right)+\sinh^2\left(c\right)\cosh\left(h\right)} \\
&= \frac{\sinh\left(h\right)\sinh\left(c\right)\left(\cosh\left(h\right)\cosh\left(c\right)-1\right)}{\cosh^2\left(h\right)\cosh\left(c\right)-\cosh\left(c\right)+\cosh^2\left(c\right)\cosh\left(h\right)-\cosh\left(h\right)} \\
&= \frac{\sinh\left(h\right)\sinh\left(c\right)}{\cosh\left(h\right)+\cosh\left(c\right)}.
\end{aligned} \tag{4}$$

Analogously,[3]

$$\begin{aligned}
&\tan\left(2\arctan\left(\tanh\left(\frac{h}{2}\right)\tanh\left(\frac{c}{2}\right)\right)\right) \\
&= \frac{2\tanh\left(\frac{h}{2}\right)\tanh\left(\frac{c}{2}\right)}{1-\tanh^2\left(\frac{h}{2}\right)\tanh^2\left(\frac{c}{2}\right)} \\
&= \frac{2\left(\cosh\left(h\right)-1\right)\left(\cosh\left(c\right)-1\right)}{\sinh\left(h\right)\sinh\left(c\right)\left(1-\left(\frac{\cosh(h)-1}{\sinh(h)}\right)^2\left(\frac{\cosh(c)-1}{\sinh(c)}\right)^2\right)} \\
&= \frac{2\sinh\left(h\right)\sinh\left(c\right)\left(\cosh\left(h\right)-1\right)\left(\cosh\left(c\right)-1\right)}{\sinh^2\left(h\right)\sinh^2\left(c\right)-\left(\cosh\left(h\right)-1\right)^2\left(\cosh\left(c\right)-1\right)^2} \\
&= \frac{\sinh\left(h\right)\sinh\left(c\right)}{\cosh\left(h\right)+\cosh\left(c\right)}.
\end{aligned} \tag{5}$$

Comparing (4) and (5), we obtain the lemma. ☐

---

[2] We essentially meant $\cosh^2(x)-\sinh^2(x)=1$ and $\tanh(2x)=(\cosh(x)-1)/\sinh(x)$.

[3] We first observed that because c and h are strictly positive, $0<\tanh(h/2)\tanh(c/2)<1$, and therefore $0<2\arctan(\tanh(h/2)\tanh(c/2))<\pi/2$.

Finally, let us recall a few facts about the Poincaré upper half-plane model of the hyperbolic plane, that is, the upper half-plane $\mathbb{H} = \{z \in \mathbb{C} \mid \mathrm{Im}(z) > 0\}$ equipped with the hyperbolic metric $d_{\mathbb{H}}$ defined by

$$d_{\mathbb{H}}(z_1, z_2) = 2 \cdot \mathrm{th}^{-1}\left(\frac{|z_1 - z_2|}{|z_1 - \overline{z_2}|}\right), \qquad \text{for all } z_1, z_2 \in \mathbb{H}.$$

With respect to this metric, the isometries are either of the form

$$z \in \mathbb{H} \longmapsto \frac{az+b}{cz+d} \quad \text{such that } (a,b,c,d) \in \mathbb{R}^4 \text{ and } ad - bc = 1, \tag{6}$$

or

$$z \in \mathbb{H} \longmapsto \frac{a\bar{z}+b}{c\bar{z}+d} \quad \text{such that } (a,b,c,d) \in \mathbb{R}^4 \text{ and } ad - bc = -1. \tag{7}$$

Furthermore, the geodesic lines are represented by the half-circles with the centers situated on the real axis or by the vertical lines. The hypercycles to a given geodesic line are the circular arcs or the half-straight lines that share the same endpoints with the geodesic line. These two kinds of hypercycles are elements of Figures 3 and 4 below.

3 An elementary solution

In the rest of this essay, we work in the Poincaré upper half-plane model. Let A and B be two distinct points at hyperbolic distance c. Let $\mathcal{G}$ be the geodesic line through A and B. Up to composition by an isometry (a mapping of the form (6) or (7)) we can assume that the coordinate of B is i and the coordinate of A is $e^c i$, therefore the line $\mathcal{G}$ corresponds to the positive imaginary axis. Let h be a positive real number and let $\mathcal{E}(h)$ be a hypercycle of axis $\mathcal{G}$ which is at distance h from $\mathcal{G}$. Up to an isometry of the form (7), we can assume that $\mathcal{E}(h)$ is the half-straight line $\{\alpha e^{i\theta_h} \mid \alpha > 0\}$ where $\theta_h = 2\tan^{-1}(e^h)$. For our problem, we reparametrize $\mathcal{E}(h)$, by setting for any real number t, C_t to be the point on $\mathcal{E}(h)$ of coordinate $e^{c/2+2t} \cdot e^{i\theta_h}$. According to this parametrization, the area function $A(t) = \text{Area}\,(ABC_t)$ is an even function, since the area remains invariant under reflection with respect to the perpendicular bisector of AB (such a reflection corresponds to the isometry $z \mapsto e^c/\bar{z}$). For any real number t, we set D_t to be the foot of the perpendicular from C_t to $\mathcal{G}$. The coordinate of D_t is then $e^{c/2+2t} i$. By applying Lemma 2 to the triangles AD_tC_t and BD_tC_t, we deduce

$$A(t) = 2\tan^{-1}\left(\tanh\left(t + \frac{c}{4}\right)\tanh\left(\frac{h}{2}\right)\right) - 2\tan^{-1}\left(\tanh\left(t - \frac{c}{4}\right)\tanh\left(\frac{h}{2}\right)\right). \tag{8}$$

For more details we refer to Figure 3. From relation (8), or even from the geometric definition of area (Formula (1) above), we deduce

$$\lim_{t\to-\infty} A(t) = 0 = \lim_{t\to+\infty} A(t). \tag{9}$$

We then have a smooth even function which vanishes at infinity. Hence, for solving our extremal problem and then proving the theorem stated above (or see Theorem 3 below), we just have to prove that $t = 0$ is the unique critical point of the area function.

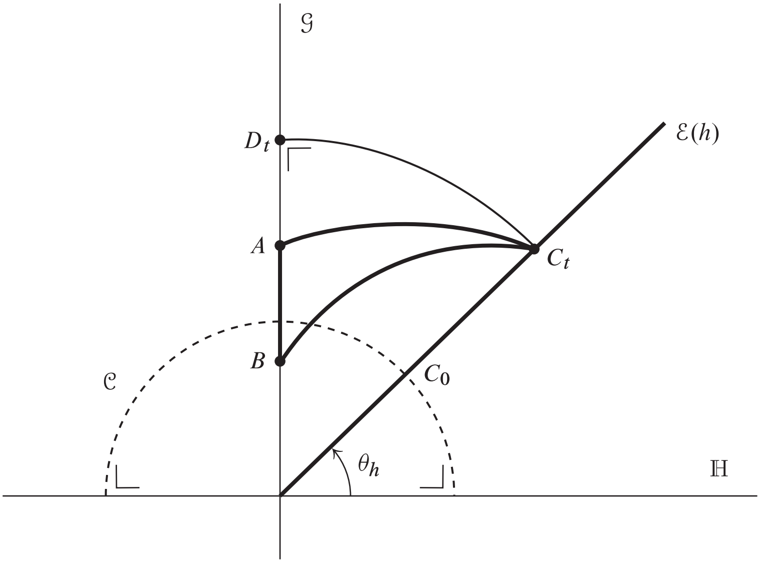


Figure 3. A geodesic line $\mathcal{G}$ with two points A and B and an equidistant curve $\mathcal{E}(h)$ at distance h from $\mathcal{G}$. Any point C_t on $\mathcal{E}(h)$ defines two right-angled triangles AD_tC_t and BD_tC_t. The perpendicular bisector of the line segment AB is denoted by $\mathcal{C}$.

We have

$$A'(t) = \frac{2\tanh\left(\frac{h}{2}\right)\left(1-\tanh^2\left(t+\frac{c}{4}\right)\right)}{1+\tanh^2\left(\frac{h}{2}\right)\tanh^2\left(t+\frac{c}{4}\right)} - \frac{2\tanh\left(\frac{h}{2}\right)\left(1-\tanh^2\left(t-\frac{c}{4}\right)\right)}{1+\tanh^2\left(\frac{h}{2}\right)\tanh^2\left(t-\frac{c}{4}\right)},$$

for all $t \in \mathbb{R}$, and then by setting

$$g(t) = \frac{2\tanh\left(\frac{h}{2}\right)}{\left(1+\tanh^2\left(\frac{h}{2}\right)\tanh^2\left(t+\frac{c}{4}\right)\right)\left(1+\tanh^2\left(\frac{h}{2}\right)\tanh^2\left(t-\frac{c}{4}\right)\right)}, \tag{10}$$

we have

$$\begin{aligned}\frac{A'(t)}{g(t)} &= \left(1-\tanh^2\left(t+\frac{c}{4}\right)\right)\left(1+\tanh^2\left(\frac{h}{2}\right)\tanh^2\left(t-\frac{c}{4}\right)\right)\\ &\quad -\left(1-\tanh^2\left(t-\frac{c}{4}\right)\right)\left(1+\tanh^2\left(\frac{h}{2}\right)\tanh^2\left(t+\frac{c}{4}\right)\right) \qquad (11)\\ &= \left(1+\tanh^2\left(\frac{h}{2}\right)\right)\left(\tanh^2\left(t-\frac{c}{4}\right)-\tanh^2\left(t+\frac{c}{4}\right)\right).\end{aligned}$$

Finally from (10) and (11), we deduce that

$$A'(t)=0 \iff t=0,$$

and thus we have proved

Theorem 3. *Let* C_0 *be the point on* $\mathcal{E}(h)$ *for which the foot of the altitude from* C_0 *is the midpoint of* AB. *Then*

$$\text{Area}(ABC) \leq \text{Area}(ABC_0) \quad \textit{for all } C \in \mathcal{E}(h).$$

Although this proof is elementary, it uses some computations, so the natural question is about the existence of a proof which is entirely geometric. The next section will show that such a proof exists and uses the notion of hyperbolic Lexell curves.

4 The use of the Lexell curve in hyperbolic geometry

Let us recall the definition of a *hyperbolic Lexell curve*. We fix a hyperbolic segment AB and we pick a third point C which is not on the geodesic containing AB. The question is to find the locus of points that define a triangle of base AB and with the same area as ABC. Such a problem in spherical geometry is called the *Lexell problem* and was solved by Lexell in [5]. In hyperbolic geometry, A'Campo and Papadopoulos solved in [1] (Theorem 5.11) that problem and the solution is as follows. We consider the geodesic line which passes through the midpoints of AC and BC. The hypercycle which passes through C and of axis this geodesic line is the solution of the Lexell problem in hyperbolic geometry. We also refer to [7] for further investigations on Lexell curves in the hyperbolic plane and to the essay [6] by Pambuccian in this volume, where the author generalizes the notion of Lexell curve in weaker geometries, namely, the *Bachmann non-elliptic* metric planes.

Let us turn back to our problem. Let A and B be two fixed points in the hyperbolic plane which lie on the geodesic line $\mathcal{G}$, and let h a positive real number. We denote by C one of the two points at distance h from the segment AB such that ABC is isosceles. We still use the notation $\mathcal{E}(h)$ for the hypercycle of axis $\mathcal{G}$ which passes through C. Up to an isometry, we can assume that we are in the situation of Figure 4, that is, the segment AB lies on a geodesic which is a half-circle and whose center is the origin and the point C lies on the imaginary axis. We then take a point D which belongs to the altitude from C to AB. It defines an isosceles triangle ABD whose area is strictly smaller than that of ABC. Furthermore, the corresponding Lexell curve intersects $\mathcal{E}(h)$ at two points. By moving along the altitude from C to AB and dealing with the corresponding Lexell curve, we define a natural surjective mapping from the hypercycle $\mathcal{E}(h)$ onto the segment that connects C and the midpoint of AB. In conclusion, we deduce that the point for which the area is maximal is the point C and also that we can have triangles of same base and altitude of area as small as possible. We can refer to Figure 4 for the construction of two Lexell curves in the Poincaré upper half-plane.

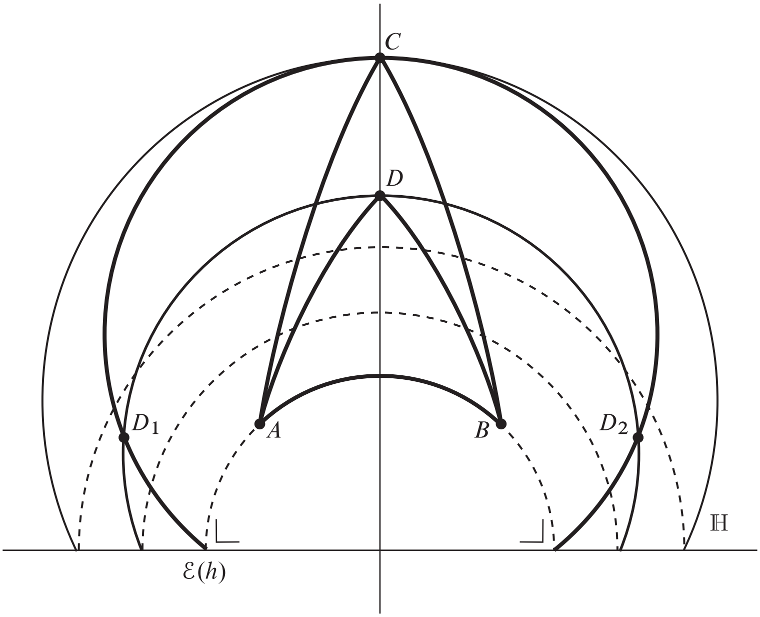


Figure 4. Two isosceles triangles ABC and ABD and their corresponding Lexell curves. The curve $\mathcal{E}(h)$ is the hypercycle passing through C of axis the geodesic line through AB. The points D_1 and D_2 are two points on $\mathcal{E}(h)$ for which Area$(ABD_1) =$ Area$(ABD_2) =$ Area(ABD).

Acknowledgements. The authors would like to express their sincere gratitude to Athanase Papadopoulos for giving them the opportunity for writing this essay and especially for his patience, kindness, and constructive remarks. They also thanks Himalaya Senapati for his constructive remarks.

The first author acknowledges support from U.S. National Science Foundation grants DMS 1107452, 1107263, 1107367 "RNMS: GEometric structures And Representation varietie" (the GEAR Network). The first author was also partially supported by the JSPS Short-term program and acknowledges the hospitality of Osaka University where some part of this work was done.

References

[1] N. A'Campo and A. Papadopoulos, "Notes on non-Euclidean geometry," in *Strasbourg master class on geometry* (A. Papadopoulos, ed.), IRMA Lectures in Mathematics and Theoretical Physics 18, European Mathematical Society (EMS), Zürich, 2012, 1–182.
MR 2931886 Zbl 1253.51001 ☞ 48, 53

[2] N. A'Campo and A. Papadopoulos, "Area in non-Euclidean geometry," in *Eighteen essays in non-Euclidean geometry* (V. Alberge and A. Papadopoulos, eds.), Chapter 1, 3–25.
☞ 49

[3] J.-M. de Tilly, "Études de mécanique abstraite," *Mémoires couronnés et autres mémoires publiés par l'Académie royale de Belgique* 21 (1870), 1–99.
☞ 49

[4] E. Frenkel and W. Su, "The area formula for hyperbolic triangles," in *Eighteen essays in non-Euclidean geometry* (V. Alberge and A. Papadopoulos, eds.), European Mathematical Society (EMS), Zürich, 2019, Chapter 2, 27–46.
☞ 48

[5] A. J. Lexell, "Solutio problematis geometrici ex doctrina sphaericorum," *Acta Academiae Scientarum Imperialis Petropolotinae* 1 (1781), 112–126.
☞ 53

[6] V. Pambuccian, "A theorem on equiareal triangles with a fixed base," in *Eighteen essays in non-Euclidean geometry* (V. Alberge and A. Papadopoulos, eds.), European Mathematical Society (EMS), Zürich, 2019, Chapter 18, 427–437.
☞ 53

[7] A. Papadopoulos and W. Su, “On hyperbolic analogues of some classical theorems in spherical geometry,” in *Hyperbolic geometry and geometric group theory* (K. Fujiwara, K. Ohshika and S. Kojima, eds.), Advanced Studies of Pure Mathematics 73, Mathematical Society of Japan, Tokyo, 2017, 225–253.
MR 3728500 ☞ 53

[8] F. T. von Schubert, “Problematis cuiusdam sphaerici solutio,” *Nova acta Academiae scientiarum imperialis petropolitanae* 4 (1786), 89–94.
☞ 47

[9] D. Slutskiy, “De Tilly’s mechanical view on hyperbolic and spherical geometries,” in *Eighteen essays in non-Euclidean geometry* (V. Alberge and A. Papadopoulos, eds.), European Mathematical Society (EMS), Zürich, 2019, Chapter 7, 93–111.
☞ 49

Chapter 4

On a theorem of Lambert: medians in spherical and hyperbolic geometries

Himalaya Senapati

Contents

1. Introduction 57
2. Ceva's theorem 58
3. Medians of spherical and hyperbolic triangles 59
4. A consequence of a theorem of Menelaus 63

References 64

1 Introduction

In this chapter, we focus on medians in spherical and hyperbolic triangles. We first state Ceva's theorem in a form that is valid in all three geometries, Euclidean, spherical and hyperbolic. As a corollary, we prove that in any of these geometries, the three medians in any triangle are concurrent. We then recall a result of Lambert stating that given an equilateral triangle ABC with medians AA', BB', and CC' intersecting at O, we have

$$OA' = \frac{1}{3}AA', \quad OA' > \frac{1}{3}AA', \quad OA' < \frac{1}{3}AA'$$

in Euclidean, spherical and hyperbolic geometry respectively [1] (see also [2] for further details). Next, we derive a relation for medians in an arbitrary triangle. More precisely, let ABC be a triangle whose three medians intersect at a point O, and A' the midpoint of BC. Then,

$$\frac{\sin AO}{\sin OA'} = 2\cos\left(\frac{BC}{2}\right) \quad \text{and} \quad \frac{\sinh AO}{\sinh OA'} = 2\cosh\left(\frac{BC}{2}\right)$$

in spherical and hyperbolic geometries respectively. We prove that in spherical geometry, when $BC > \frac{2\pi}{3}$ and $AO < \frac{\pi}{2}$, $AO < 2\cos\frac{BC}{2} \times OA'$, implying

$$OA' > \frac{AA'}{1 + 2\cos\left(\frac{BC}{2}\right)}.$$

We prove that in hyperbolic geometry $AO < 2\cosh\frac{BC}{2}$, implying

$$OA' > \frac{AA'}{1 + 2\cosh\left(\frac{BC}{2}\right)}.$$

We conclude this chapter by stating a theorem due to Menelaus where he studies a geodesic intersecting three sides of a spherical triangle. As a corollary, we prove that in a spherical triangle ABC, the midpoints of AB, AC, BC', and CB' are collinear where B' and C' are antipodal points of B and C respectively.

2 Ceva's theorem

We state a theorem due to Ceva on concurrent lines in Euclidean geometry and its analogues in spherical and hyperbolic geometries (see [2] for a proof).

Theorem 2.1 (Ceva's theorem). *Let ABC be a triangle. Let AA', BB', and CC' be concurrent lines with A', B', and C' lying on BC, CA, and AB. Then, in Euclidean geometry, we have*

$$\frac{CA'}{A'B} \times \frac{BC'}{C'A} \times \frac{AB'}{B'C} = 1,$$

while in spherical and hyperbolic geometries we have respectively

$$\frac{\sin(CA')}{\sin(A'B)} \times \frac{\sin(BC')}{\sin(C'A)} \times \frac{\sin(AB')}{\sin(B'C)} = 1$$

and

$$\frac{\sinh(CA')}{\sinh(A'B)} \times \frac{\sinh(BC')}{\sinh(C'A)} \times \frac{\sinh(AB')}{\sinh(B'C)} = 1.$$

We can restate the above theorem in a form that is true in the three geometries.

Theorem 2.2. *Let ABC be a triangle and let AA', BB', and CC' be concurrent lines. Then, in all three geometries, Euclidean, spherical and hyperbolic, we have*

$$\frac{\sin(\widehat{CAA'})}{\sin(\widehat{A'AB})} \times \frac{\sin(\widehat{ABB'})}{\sin(\widehat{B'BC})} \times \frac{\sin(\widehat{BCC'})}{\sin(\widehat{C'CA})} = 1.$$

Proof. In the following, we give a proof of this theorem in spherical geometry. The proof in the case of Euclidean and hyperbolic geometries is very similar with an appropriate modification to the law of sines. Let the concurrent lines meet at the point O. Applying the law of sines to the triangle AOB, we find

$$\frac{\sin(OA)}{\sin(\widehat{OBA})} = \frac{\sin(OB)}{\sin(\widehat{OAB})} \implies \frac{\sin(\widehat{OAB})}{\sin(\widehat{OBA})} = \frac{\sin(OB)}{\sin(OA)}.$$

Similarly, using the law of sines in triangles BOC and COA, we find

$$\frac{\sin(\widehat{OBC})}{\sin(\widehat{OCB})} = \frac{\sin(OC)}{\sin(OB)} \quad \text{and} \quad \frac{\sin(\widehat{OCA})}{\sin(\widehat{OAC})} = \frac{\sin(OA)}{\sin(OC)}.$$

Multiplying the above three equations, we find

$$\frac{\sin(\widehat{CAA'})}{\sin(\widehat{A'AB})} \times \frac{\sin(\widehat{ABB'})}{\sin(\widehat{B'BC})} \times \frac{\sin(\widehat{BCC'})}{\sin(\widehat{C'CA})} = \frac{\sin(OB)}{\sin(OA)} \times \frac{\sin(OC)}{\sin(OB)} \times \frac{\sin(OA)}{\sin(OC)} = 1.$$

□

Corollary 2.3. *The medians of a triangle are concurrent in Euclidean, spherical and hyperbolic geometries.*

Proof. We prove it for a spherical triangle. The other proofs are similar. Let ABC be a spherical triangle. Since $AB = BC = CA = \pi$ is not possible, we take without loss of generality, $BC \neq \pi$. Let BB' and CC' be medians and let them intersect at O. Extend the great circle arc AO to intersect BC at A'. Since $B'C = B'A$ and $C'A = C'B$, by Ceva's theorem, we must have $\sin A'B = \sin CA'$ or equivalently, $A'B = CA'$ since $BC \neq \pi$. □

3 Medians of spherical and hyperbolic triangles

We now recall a classical theorem due to Lambert on lengths of medians of equilateral triangles. See [2] for a proof and further details.

Theorem 3.1 (Lambert's theorem [1]). *Let ABC be an equilateral triangle and let the medians AA', BB', and CC' intersect at O. We have $OA' = \frac{1}{3}AA'$, $OA' > \frac{1}{3}AA'$, and $OA' < \frac{1}{3}AA'$ for Euclidean, spherical and hyperbolic geometries.*

In the following, we deduce a precise relation between AO and OA'. We then prove an appropriate inequality between OA' and AA' for arbitrary triangles in spherical and hyperbolic geometries.

Theorem 3.2. *Let ABC be a triangle with the midpoints of AB, BC, and CA denoted by C', A', and B'. Let the medians AA', BB', and CC' intersect at O. Then, we have*

$$\frac{\sin(AO)}{\sin(OA')} = 2\cos\left(\frac{BC}{2}\right) \quad \textit{and} \quad \frac{\sinh(AO)}{\sinh(OA')} = 2\cosh\left(\frac{BC}{2}\right)$$

in spherical and hyperbolic geometries respectively.

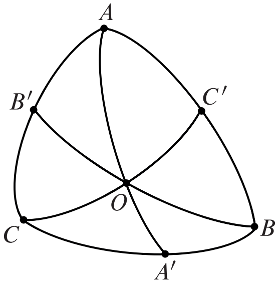


Figure 1. Spherical triangle ABC with medians AA', BB', and CC'

Proof. We begin by proving the theorem in spherical geometry. Let ABC be a spherical triangle and let A', B', and C' be the midpoints of BC, CA, and AB (see Figure 1). Let the medians AA', BB', and CC' intersect at O. Using the law of sines in the triangles AOB and $A'OB$, we have

$$\frac{\sin(\widehat{ABB'})}{\sin(AO)} = \frac{\sin(\widehat{BAA'})}{\sin(BO)} \quad \text{and} \quad \frac{\sin(\widehat{CBB'})}{\sin(OA')} = \frac{\sin(\widehat{AA'B})}{\sin(BO)}.$$

Dividing side by side the two equations, we find

$$\frac{\sin(OA')}{\sin(AO)} \times \frac{\sin(\widehat{ABB'})}{\sin(\widehat{CBB'})} = \frac{\sin(\widehat{BAA'})}{\sin(\widehat{AA'B})}. \tag{1}$$

In the triangle $AA'B$, we have

$$\frac{\sin(AB)}{\sin(\widehat{AA'B})} = \frac{\sin(A'B)}{\sin(\widehat{BAA'})} \implies \frac{\sin(\widehat{BAA'})}{\sin(\widehat{AA'B})} = \frac{\sin\left(\frac{1}{2}BC\right)}{\sin(AB)}. \tag{2}$$

Using the law of sines in the triangles $BB'C$ and $BB'A$, we have

$$\frac{\sin(\widehat{CBB'})}{\sin(B'C)} = \frac{\sin(\widehat{BB'C})}{\sin(BC)} \quad \text{and} \quad \frac{\sin(AB')}{\sin(\widehat{ABB'})} = \frac{\sin(AB)}{\sin(\widehat{BB'A})}.$$

Multiplying the two equations and using $B'C = AB'$ and $\sin\widehat{BB'C} = \sin(\widehat{BB'A})$, we get

$$\frac{\sin(\widehat{CBB'})}{\sin(\widehat{ABB'})} = \frac{\sin(AB)}{\sin(BC)}. \tag{3}$$

Substituting (2) and (3) in (1), we find

$$\frac{\sin(OA')}{\sin(AO)} = \frac{\sin\left(\frac{1}{2}BC\right)}{\sin(AB)} \times \frac{\sin(AB)}{\sin(BC)} = \frac{1}{2\cos\left(\frac{1}{2}BC\right)} \tag{4}$$

thus proving the theorem in spherical geometry.

In the hyperbolic case, the law of sines in a hyperbolic triangle ABC gets modified to

$$\frac{\sin(\widehat{BAC})}{\sinh(BC)} = \frac{\sin(\widehat{CBA})}{\sinh(CA)} = \frac{\sin(\widehat{ACB})}{\sinh(AB)}$$

where the sine function on the edges has been replaced by the hyperbolic sine function. Thus, proceeding in a similar manner, equation (4) becomes

$$\frac{\sinh(AO)}{\sinh(OA')} = 2\cosh\left(\frac{1}{2}BC\right). \qquad \square$$

We now prove two lemmas that will be used in proving Corollary 3.5.

Lemma 3.3. *If* $0 < b < a < \pi$, *then* $a\sin(b) > b\sin(a)$.

Proof. Setting $k = \frac{b}{a}$, we have

$$\sin(b) - \frac{b}{a}\sin(a) = k\int_0^a (\cos(kx) - \cos(x))dx.$$

Since $k < 1$ and $0 < kx < x < \pi$, we have $\cos(kx) > \cos(x)$. This shows that the right-hand side of the above is positive proving that $\sin(b) > \frac{b}{a}\sin(a)$ or, equivalently, $a\sin(b) > b\sin(a)$. $\square$

Lemma 3.4. *If* $a > 0$ *and* $k > 1$, *then* $k\sinh(a) < \sinh(ak)$.

Proof. We note that

$$\sinh(ak) - k\sinh(a) = k\int_0^a (\cosh(kx) - \cosh(x))\,dx.$$

Since $k > 1$ and $x > 0$, $\cosh(kx) > \cosh(x)$ and thus the right-hand side is positive. This proves that $\sinh(ak) > k\sinh(a)$. $\square$

Theorem 3.5. *Let ABC be a triangle with medians AA', BB', and CC' intersecting at O. In spherical geometry, when AO is less than a quarter of a circle, we have*

$$AO \begin{cases} < 2\cos\left(\frac{1}{2}BC\right) \times OA' & \textit{if } BC > \frac{2\pi}{3}, \\ > 2\cos\left(\frac{1}{2}BC\right) \times OA' & \textit{if } BC < \frac{2\pi}{3} \textit{ and } 2\cos\left(\frac{1}{2}BC\right) \times OA' < \frac{\pi}{2}. \end{cases}$$

In hyperbolic geometry, we have

$$AO < 2\cosh\left(\frac{1}{2}BC\right) \times OA'.$$

Proof. We start by proving the theorem in spherical geometry. Theorem 3.2 implies

$$\sin(AO) = 2\cos\left(\frac{1}{2}BC\right) \times \sin(OA').$$

When $2\cos\left(\frac{1}{2}BC\right) < 1$ or $2\cos\left(\frac{1}{2}BC\right) \times OA' < OA'$, by Lemma 3.3,

$$OA' \times \sin\left(2\cos\left(\frac{1}{2}BC\right) \times OA'\right) > 2\cos\left(\frac{1}{2}BC\right) \times OA' \times \sin(OA')$$

implying $\sin\left(2\cos\left(\frac{1}{2}BC\right) \times OA'\right) > \sin AO$. Since $0 < OA' < \pi$, when $AO < \frac{\pi}{2}$, this implies

$$AO < 2\cos\left(\frac{1}{2}BC\right) \times OA'.$$

Similarly, when $2\cos\left(\frac{1}{2}BC\right) > 1$, Lemma 3.3 implies

$$\sin\left(2\cos\left(\frac{1}{2}BC\right) \times OA'\right) < \sin(AO).$$

When $AO < \frac{\pi}{2}$, this leads to

$$AO > 2\cos\left(\frac{1}{2}BC\right) \times OA' \quad \text{if } 2\cos\left(\frac{1}{2}BC\right) \times OA' < \frac{\pi}{2}.$$

We conclude the proof in the spherical case by observing that if $BC < \frac{2\pi}{3}$, then $2\cos\left(\frac{1}{2}BC\right) > 2\cos\left(\frac{\pi}{3}\right) = 1$ while if $\frac{2\pi}{3} < BC < \pi$, then $2\cos\left(\frac{1}{2}BC\right) < 1$.

In hyperbolic geometry, $2\cosh\left(\frac{1}{2}BC\right) > 1$. By Theorem 3.2 and Lemma 3.4,

$$\sinh(AO) = 2\cosh\left(\frac{1}{2}BC\right)\sinh(OA') < \sinh\left(2\cosh\left(\frac{1}{2}BC\right) \times A'O\right).$$

This implies $AO < 2\cosh\left(\frac{1}{2}BC\right) \times A'O$ proving the theorem in hyperbolic geometry. ☐

4 A consequence of a theorem of Menelaus

We state below a theorem due to Menelaus [3] that will be used to show collinearity of midpoints of two adjacent sides of a spherical triangle and two antipodal points on the extended third side.

Theorem 4.1. (*Spherics*, Proposition 66). *Let ABC be a spherical triangle and let a geodesic intersect AB at D, BC at E, and CA at F. Then,*

$$\frac{\sin(AD)}{\sin(BD)} \times \frac{\sin(BE)}{\sin(CE)} \times \frac{\sin(CF)}{\sin(AF)} = 1.$$

Corollary 4.2. *Let ABC be an arbitrary spherical triangle. Complete the great circle BC and denote the antipodal points of B and C by B' and C'. Also, denote the midpoint of BC' by X and its antipodal point by Y. Consider a great circle passing through X and Y other than the one passing through B and C. Let it intersect AB at E and AC at F (see Figure 2). Then,*

$$\frac{\sin(AE)}{\sin(BE)} = \frac{\sin(AF)}{\sin(CF)}.$$

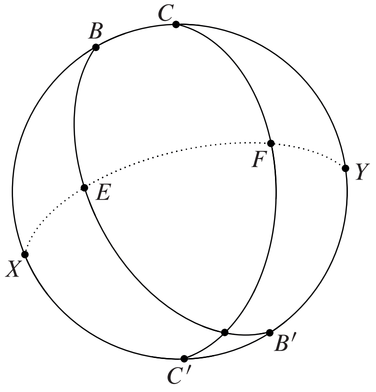


Figure 2. ABC is a spherical triangle. B', C', and Y are antipodal to B, C, and X where $BX = XC'$.

Proof. By Theorem 4.1,

$$\frac{\sin(AE)}{\sin(BE)} \times \frac{\sin(BX)}{\sin(CX)} \times \frac{\sin(CF)}{\sin(AF)} = 1.$$

Since $BX = XC'$,

$$\pi = CC' = CX + XC' = CX + BX \implies \sin(CX) = \sin(BX).$$

Hence,

$$\frac{\sin(AE)}{\sin(BE)} = \frac{\sin(AF)}{\sin(CF)}. \qquad \square$$

Corollary 4.3. *The great circle through the midpoints of AB and AC passes through X and Y.*

Proof. Let E be the midpoint of AB and let the great circle XEY intersect AC at F. By the above corollary,

$$\frac{\sin(AF)}{\sin(FC)} = \frac{\sin(AE)}{\sin(EB)} = 1,$$

implying $\sin AF = \sin FC$. However, since $0 < AC < \pi$, we have $AF = FC$. $\square$

Acknowledgements. The author would like to thank Athanase Papadopoulos and Vincent Alberge for their critical comments and valuable suggestions. The author acknowledges the hospitality of ICTS/Prog-ggd/2017/11 where a part of this work was done.

References

[1] J. H. Lambert, “Theorie der Parallellinien,” in *Die Theorie der Parallellinien von Euklid bis auf Gauss, eine Urkundensammlung zur Vorgeschichte der nicht-euklidischen Geometrie* (P. Stäckel and F. Engel, eds.), B. G. Teubner, Leipzig 1895. French translation by A. Papadopoulos and G. Théret, *La théorie des lignes parallèles de Johann Heinrich Lambert,* Sciences dans l’Histoire, Blanchard, Paris 2014.
☞ 57, 59

[2] A. Papadopoulos and W. Su, “On hyperbolic analogues of some classical theorems in spherical geometry,” in *Hyperbolic geometry and geometric group theory* (K. Fujiwara, K. Ohshika, and S. Kojima, eds.), Advanced Studies of Pure Mathematics 73, Mathematical Society of Japan, Tokyo, 2017, 225–253.
MR 3728500 ☞ 57, 58, 59

[3] R. Rashed and A. Papadopoulos, *Menelaus' spherics*: *early translation and al-Māhānī/al-Harawī's version,* Critical edition with English translation and historical and mathematical commentaries, Series Scientia Graeco-Arabica 21, de Gruyter, Berlin, 2017.
Zbl 06908518 ☞ 63

Chapter 5

Inscribing a triangle in a circle in spherical geometry

Himalaya Senapati

Contents

1. Introduction 67
2. Lagrange's solution 68
3. Constructibility on the sphere 72
4. Extension of the problem to spherical geometry 75

References 78

1 Introduction

We consider the geometrical problem of constructing a triangle using a straightedge and compass so that its three vertices lie on a given circle and its three sides when produced pass through three given points. Pappus of Alexandria formulated and solved this problem in the special case where the three given points are aligned. This is Proposition 117 of Book VII of the *Collection* [1]. Lagrange provided an algebraic solution, proving that the vertices of the triangle are constructible points. He communicated his solution to Jean de Castillon, who reproduced it in the first section of his paper "Sur une nouvelle propriété des sections coniques" (On a new property on conic sections) [2]. In the rest of his this paper, Castillon proceeded to show that the vertices of the triangle remain constructible when the circle is replaced by an ellipse or a hyperbola. Later, the solution due to Lagrange was published in his *Oeuvres complètes* under the title "Solution algébrique d'un problème de géométrie" (Algebraic solution of a geometry problem) [3]. Euler gave another solution and commented on the case where the given circle and the three given points are on a sphere [4]. Interestingly, this article is followed by one due to Nicolas Fuss, a young collaborator of Euler, who similar to Pappus, studies the special case where the three given points lie on a line [5]. Eugène-Charles Catalan also studied this problem in his book *Théorèmes et problèmes de géométrie élémentaire* (Theorems and problems of elementary geometry) [6] where he says that such a problem is known as the "Castillon problem" (see the footnote at p. 222 of [6]).

In this chapter, we first recall the algebraic solution due to Lagrange. The idea behind such a solution is to prove that the vertices of the triangle are constructible. More precisely, it is to prove that the coordinates of the vertices are roots of some quadratic polynomials with constructible coefficients. Let us recall the notion of constructibility that is relevant here. Let $\mathcal{S}$ be a given set of points in the Euclidean plane with cardinality at least 2.[1] We say that a circle is *constructible from* $\mathcal{S}$ if its center is an element of $\mathcal{S}$ and its radius is the distance between two elements of $\mathcal{S}$. We say that a line is constructible from $\mathcal{S}$ if it passes through two elements of $\mathcal{S}$. We say that a point in the Euclidean plane is *constructible in one step* from $\mathcal{S}$ if it is an intersection point of two constructible lines, one constructible line and a constructible circle, or two constructible circles. We then say that a point P of the Euclidean plane is *constructible* from $\mathcal{S}$ if there exists a finite sequence of points $P_1, P_2, \ldots, P_n = P$ such that P_1 is constructible in one step from $\mathcal{S}$ and for any i $(2 \leq i \leq n)$, the point P_i is constructible in one step from $\mathcal{S} \cup \{P_1, \ldots, P_{i-1}\}$. In particular, let us denote two points from $\mathcal{S}$ by O and I and set the distance from O to I as 1. We can now introduce the Cartesian coordinate system in the Euclidean plane where the coordinates of O and I are $(0, 0)$ and $(1, 0)$ respectively. We say that a number x is *constructible from* $\mathcal{S}$ if it is a coordinate of a constructible point from $\mathcal{S}$. It is elementary to prove that the set of constructible numbers from $\mathcal{S}$ forms a field such that the square root of any non-negative constructible number from $\mathcal{S}$ is still a constructible number from $\mathcal{S}$. In particular, a root of a quadratic polynomial with constructible (from $\mathcal{S}$) coefficients is constructible.

In the next section, we recall Lagrange's solution. We then study the problem where the given circle and the three given points are on a sphere and give a solution in the lineage of Lagrange.

2 Lagrange's solution

The problem is to construct a triangle MNP inscribed in a given circle $\mathcal{C}$ of radius r so that the three edges of the triangle NM, MP, and PN when produced pass through three given points A, B, and C respectively (see Figure 1). This construction should not require more than the given objects and a straightedge and a compass. To be more precise, the set we are dealing with consists of six points, the two points O and I defined above, the point of coordinates $(r, 0)$, and the three points A, B, and C. Let us just add that the center of the given circle can be taken as the origin O.

---

[1] We are using a modern way to formulate constructibility. Needless to say, Lagrange did not use this language of set theory or of fields.

We notice that the vertices M, N, and P are constructible if the angles $x = \widehat{AOM}$, $y = \widehat{AON}$ and $z = \widehat{AOP}$ are constructible, where an angle is said to be constructible if its sine (or equivalently cosine) is constructible. We set the lengths of segments

$$OA = a, \quad OB = b, \quad OC = c.$$

We denote the angles $\widehat{AOB}$ and $\widehat{AOC}$ by m and n. Here, 1, a, b, c, $\cos(m)$, $\cos(n)$, and r are constructible. In the triangle AON, we have

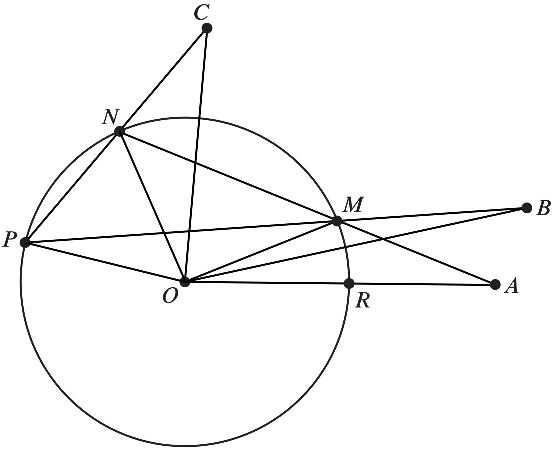


Figure 1. An inscribed triangle MNP such that NM, MP, and PN when produced pass through A, B, and C respectively

$$\widehat{ONA} = \frac{1}{2}(\pi - \widehat{NOM}) = \frac{\pi}{2} - \frac{y-x}{2}$$

and

$$\widehat{OAN} = \pi - \widehat{AON} - \widehat{ONA} = \frac{\pi}{2} - \frac{y+x}{2}.$$

Using the sine law in the triangle AON, we have

$$\frac{OA}{ON} = \frac{\sin(\widehat{ONA})}{\sin(\widehat{OAN})} \implies \frac{a}{r} = \frac{\cos\left(\frac{y-x}{2}\right)}{\cos\left(\frac{y+x}{2}\right)}.$$

This reduces to

$$\begin{aligned} &a\Big(\cos\left(\frac{y}{2}\right)\cos\left(\frac{x}{2}\right)-\sin\left(\frac{y}{2}\right)\sin\left(\frac{x}{2}\right)\Big)\\ &\qquad = r\Big(\cos\left(\frac{y}{2}\right)\cos\left(\frac{x}{2}\right)+\sin\left(\frac{y}{2}\right)\sin\left(\frac{x}{2}\right)\Big)\\ &\implies (a-r)\cos\left(\frac{y}{2}\right)\cos\left(\frac{x}{2}\right) = (a+r)\sin\left(\frac{y}{2}\right)\sin\left(\frac{x}{2}\right)\\ &\implies \tan\left(\frac{x}{2}\right)\tan\left(\frac{y}{2}\right) = \frac{a-r}{a+r}. \end{aligned} \tag{1}$$

In the triangle BOP, we find

$$\widehat{OPB} = \frac{\pi}{2} - \frac{z-x}{2} \quad \text{and} \quad \widehat{OBP} = \frac{\pi}{2} + m - \frac{z+x}{2}.$$

Using the sine law in the triangle BOP, we have

$$\frac{BO}{PO} = \frac{\sin(\widehat{OPB})}{\sin(\widehat{OBP})} \implies \frac{b}{r} = \frac{\cos\left(\frac{(z-m)-(x-m)}{2}\right)}{\cos\left(\frac{(z-m)+(x-m)}{2}\right)}$$

which reduces to

$$\tan\frac{x-m}{2}\tan\frac{z-m}{2} = \frac{b-r}{b+r}. \tag{2}$$

Similarly, in the triangle COP, we have

$$\widehat{OPC} = \frac{\pi}{2} - \frac{z-y}{2} \quad \text{and} \quad \widehat{OCP} = \frac{\pi}{2} + n - \frac{z+y}{2}.$$

This leads to

$$\tan\left(\frac{y-n}{2}\right)\tan\left(\frac{z-n}{2}\right) = \frac{c-r}{c+r}. \tag{3}$$

The goal is now to determine the unknowns x, y, and z from equations (1)–(3). For simplicity, let us set

$$\tan\left(\frac{x}{2}\right) = s, \quad \tan\left(\frac{y}{2}\right) = t, \quad \tan\left(\frac{z}{2}\right) = u, \quad \tan\left(\frac{m}{2}\right) = p, \quad \tan\left(\frac{n}{2}\right) = q,$$

$$\frac{a-r}{a+r} = \alpha, \quad \frac{b-r}{b+r} = \beta, \quad \frac{c-r}{c+r} = \gamma.$$

Here, α, β, γ, p, and q are constructible since a, b, c, m, n, and r are constructible. Equations (1)–(3) now become

$$s \times t = \alpha, \quad \frac{s-p}{1+ps} \times \frac{u-p}{1+pu} = \beta, \quad \frac{t-q}{1+qt} \times \frac{u-q}{1+qu} = \gamma.$$

The first and the second give

$$t = \frac{\alpha}{s} \quad \text{and} \quad u = \frac{\beta(1+ps)+p(s-p)}{(s-p)-\beta p(1+ps)}. \tag{4}$$

Substituting the first and the second in the third gives

$$\frac{\alpha - qs}{\alpha q + s} \times \frac{p(-p+q)+\beta(1+pq)+(p-q+\beta p(1+pq))s}{-(1+\beta)p+\beta q-p^2q+(1-\beta p^2+(1+\beta)pq)s} = \gamma. \tag{5}$$

For conciseness, we set

$$F = p(-p+q)+\beta(1+pq) \quad \text{and} \quad G = p-q+\beta p(1+pq).$$

We also set

$$H = -(1+\beta)p+\beta q-p^2q \quad \text{and} \quad K = 1-\beta p^2+(1+\beta)pq.$$

Therefore, F, G, H, and K are constructible since β, p, and q are constructible. Equation (5) now becomes

$$\frac{\alpha - qs}{\alpha q + s} \times \frac{F+Gs}{H+Ks} = \gamma$$

which implies

$$(\gamma K - Gq) \times s^2 + (\gamma H - \alpha G + \alpha q(\gamma K - F)) \times s + \alpha(\gamma H q - F) = 0. \tag{6}$$

The coefficients of this quadratic equation in s are constructible, hence s is also constructible. This implies that the right-hand sides of the two equations in (4) are constructible and thus implying that t and u are constructible. Thus

$$\cos(x) = \frac{1-s^2}{1+s^2}, \quad \cos(y) = \frac{1-t^2}{1+t^2}, \quad \cos(z) = \frac{1-u^2}{1+u^2}$$

are constructible. For the existence of the vertices M, N, and P, the angles $\frac{x}{2}$, $\frac{y}{2}$, and $\frac{z}{2}$ and hence the tangents s, t, and u should be real. When s is real, (4) implies $t, u \in \mathbb{R}$. Hence we have proved that when equation (6) admits real roots, a triangle exists and is constructible so that its vertices lie on a given circle and its three edges when produced pass through three given points. When equation (6) does not admit real roots, such a triangle does not exist.

Remark 2.1. We now give an example where such a triangle does not exist. Let the three given points and the centre of the given circle be collinear such that $m = n = 0$. This implies

$$p = \tan\left(\frac{m}{2}\right) = 0 \quad \text{and} \quad q = \tan\left(\frac{n}{2}\right) = 0.$$

Substituting these values, we find $F = \beta$, $G = 0$, $H = 0$, and $K = 1$. Thus, equation (6) becomes

$$\gamma \times s^2 - \alpha\beta = 0$$

with discriminant

$$4\alpha\beta\gamma = 4 \times \frac{a-r}{a+r} \times \frac{b-r}{b+r} \times \frac{c-r}{c+r}.$$

This is negative when only one of the three points is in the interior of the circle or when all three are in the interior. Thus, by continuity, there exist approximately collinear points ($m \approx n \approx 0$, $m, n \neq 0$) with either only one of the points in the interior or all three in the interior such that the discriminant is negative. In all these cases, the triangle to be inscribed does not exist.

3 Constructibility on the sphere

We define here a notion of constructibility on the sphere. This requires the use of an auxiliary Euclidean plane. Let us assume that the sphere is of radius 1. We report on the auxiliary plane two points at distance 1. Let us denote these two points by O and I. Now, let $\mathcal{S}_+$ be a set of points on the sphere. On the plane, we draw circles centered at O of radius the chord distance between two points of the given set. We set $\mathcal{S}_0$ as the set which consists of the two points O, I and the intersection points of those circles with the half-line starting at O and passing through I. We say that a circle is *constructible* from $\mathcal{S}_+$ if its pole is a point of $\mathcal{S}_+$ and the chord distance to the pole is a constructible number from $\mathcal{S}_0$. Like for the case of constructibility in the plane, we say that a point on the sphere is *constructible in one step* from $\mathcal{S}_+$ if it is an intersection point of two constructible circles. Thus, a point P on the sphere is called constructible if there exists a finite sequence of points $P_1, P_2, \ldots, P_n = P$ such that P_1 is constructible in one step from $\mathcal{S}_+$ and for $2 \leq i \leq n$, P_i is constructible in one step from $\mathcal{S}_+ \cup \{P_1, \ldots, P_{i-1}\}$.

Following our definition of constructibility on the sphere, we state some consequences which shall be useful in Section 4.

Lemma 3.1. *A great circle on the sphere is constructible if one of its poles is constructible.*

Proof. Let C be the centre of the sphere and P the constructible pole (see Figure 2). Let Q be an arbitrary point on the great circle with pole P. This implies $PC = QC = 1$ and $\widehat{PCQ} = \frac{\pi}{2}$ so that $PQ = \sqrt{2}$. Thus the chord distance of a point on the great circle from its pole is constructible on the plane implying that the great circle is constructible □

Lemma 3.2. *A geodesic through two given points on the sphere is constructible.*

Proof. Let P and Q be the two given points (see Figure 3). We construct the two great circles whose poles are P and Q respectively. Let a point of intersection be called C'. We have the chord lengths $PC' = QC' = \sqrt{2}$. Thus the great circle whose pole is C' passes through P and Q. Since C' is constructible, geodesic through P and Q is constructible. ☐

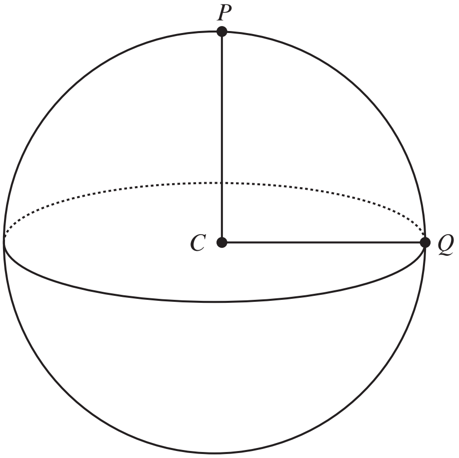


Figure 2

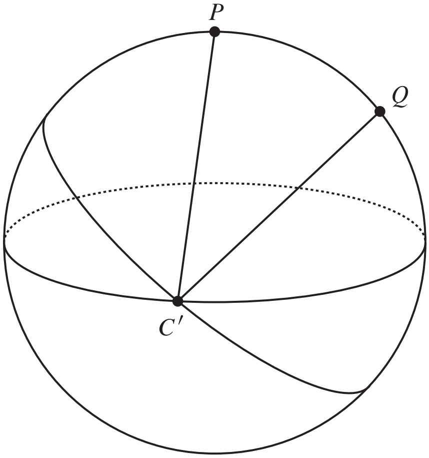


Figure 3

Lemma 3.3. *The sine of the angular separation between two constructible points on the sphere is constructible in the Euclidean plane.*

Proof. Let D and E be two given points on the sphere (see Figure 4). Construct the geodesic DE and extend it. Using a compass, find the point F on the extended geodesic so that $\operatorname{crd}(DE) = \operatorname{crd}(EF)$ where crd denotes the chord length. We have $\operatorname{crd}(DF) = 2\sin(DE)$. Since the chord distance on the sphere is constructible on the plane, $\sin(DE)$ is constructible in the Euclidean plane. ☐

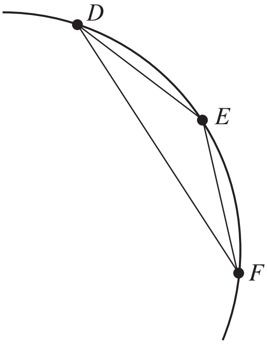


Figure 4. Determination of the sine of the angular separation between two constructible points

Lemma 3.4. *Suppose an angle θ is constructible in the Euclidean plane. Let D and E be two points on the sphere. Then, it is possible to construct a great circle arc DG such that $\widehat{EDG} = \theta$.*

Proof. Construct the great circle $\mathcal{C}_D$ with pole D. Extend the geodesic DE in the direction of E and denote the first point of intersection with this great circle by F (see Figure 5). Since θ is constructible in the Euclidean plane, $2\sin\left(\frac{\theta}{2}\right)$ is also constructible. Construct the circle $\mathcal{C}_F$ with pole F and chord distance to the pole set at $2\sin\left(\frac{\theta}{2}\right)$. Let G be a point of intersection of $\mathcal{C}_D$ and $\mathcal{C}_F$ so that $\operatorname{crd}(FG) = 2\sin\left(\frac{\theta}{2}\right)$. Thus the great circle arc FG subtends an angle θ at the centre of the sphere. Therefore, it also subtends the same angle θ at D since D is the pole of the great circle arc FG. Hence, joining DG, we have $\widehat{FDG} = \theta$. ☐

Lemma 3.5. *Suppose DE and DG are two great circle arcs on the sphere. The angle $\widehat{EDG}$ is constructible in the Euclidean plane*

Proof. The proof is very similar to that of Lemma 3.4. ☐

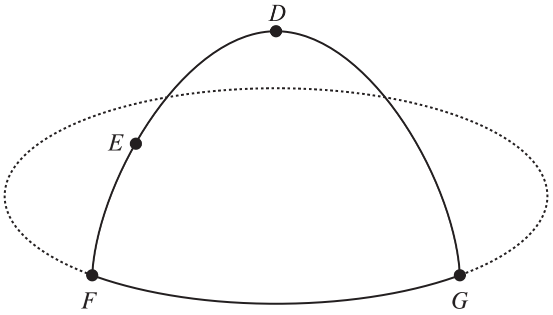


Figure 5. Given DE, constructing DG such that the angle at D is a given angle

4 Extension of the problem to spherical geometry

We now consider the extension of this construction problem to spherical geometry. Here, the problem remains to construct a triangle MNP using a straightedge and a compass whose vertices lie on a given circle $\mathcal{C}$ and whose edges when produced pass through three given points A, B, and C. However, the given circle and the three given points now lie on a sphere instead of the Euclidean plane (see Figure 6).

Let O be a pole of the circle $\mathcal{C}$. We set, as before, $OM = ON = OP = r$, $OA = a$, $OB = b$, $OC = c$, $\widehat{AOB} = m$, and $\widehat{AOC} = n$. We denote the unknown angles $\widehat{AOM}$, $\widehat{AON}$ and $\widehat{AOP}$ by x, y, and z respectively.

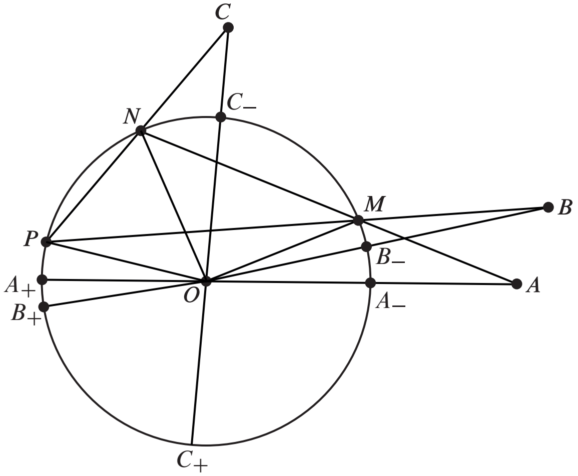


Figure 6. The circle $\mathcal{C}$ is on the surface of a sphere

By Lemma 4.2 (to be given below), the geodesic AMN intersecting the circle $\mathcal{C}$ at M and N gives the relation

$$\tan\left(\frac{x}{2}\right)\tan\left(\frac{y}{2}\right) = \frac{\sin(a-r)}{\sin(a+r)} = \alpha_+.$$

This is reminiscent of equation (1). Similarly, the geodesics BMP and CNP give identities reminiscent of equations (2) and (3):

$$\tan\left(\frac{x-m}{2}\right)\tan\left(\frac{z-m}{2}\right) = \frac{\sin(b-r)}{\sin(b+r)} = \beta_+$$

and

$$\tan\left(\frac{y-n}{2}\right)\tan\left(\frac{z-n}{2}\right) = \frac{\sin(c-r)}{\sin(c+r)} = \gamma_+.$$

As before, we set

$$\tan\left(\frac{x}{2}\right) = s, \quad \tan\left(\frac{y}{2}\right) = t, \quad \tan\left(\frac{z}{2}\right) = u, \quad \tan\left(\frac{m}{2}\right) = p, \quad \tan\left(\frac{n}{2}\right) = q.$$

Since the angles m and n are constructible on the sphere, by Lemma 3.5, they are constructible on the plane. Hence, $p = \tan\left(\frac{m}{2}\right)$ and $q = \tan\left(\frac{n}{2}\right)$ are also constructible on the plane. We also set F, G, H, and K:

$$\begin{aligned}
F &= p(-p+q) + \beta_+(1+pq),\\
G &= p - q + \beta_+ p(1+pq),\\
H &= -(1+\beta_+)p + \beta_+ q - p^2 q,\\
K &= 1 - \beta_+ p^2 + (1+\beta_+)pq.
\end{aligned}$$

With this notation, we obtain the equations

$$t = \frac{\alpha_+}{s}, \quad u = \frac{\beta_+(1+ps) + p(s-p)}{(s-p) - \beta_+ p(1+ps)}, \quad \frac{\alpha_+ - qs}{\alpha_+ q + s} \times \frac{F+Gs}{H+Ks} = \gamma_+. \tag{7}$$

It is clear that s, t, and u are constructible in the Euclidean plane if α_+, β_+, and γ_+ are constructible on the plane. Extending the great circle arcs OA, OB, and OC, and denoting the point of intersections with the circle by $A_\pm$, $B_\pm$, and $C_\pm$ as shown in Figure 6, we find $AA_\pm = a \pm r$, $BB_\pm = b \pm r$, and $CC_\pm = c \pm r$ that are constructible on the sphere. Hence, by Lemma 3.3, $\sin(a \pm r)$, $\sin(b \pm r)$, and $\sin(c \pm r)$ are constructible on the plane that implies α_+, β_+, and γ_+ and hence s, t, and u are constructible in the Euclidean plane. As shown before in the planar case, when $s \in \mathbb{R}$, the angles x, y, and z are real and are constructible in the Euclidean plane. By Lemma 3.4, we can construct geodesics OM, ON, and OP so that they make angles x, y, and z with OA so that the vertices M, N, and P are constructible. When $s \notin \mathbb{R}$, the required triangle does not exist. We have thus generalized the algebraic solution due to Lagrange to the spherical setting.

Remark 4.1. We give an example where the triangle to be inscribed does not exist. Consider the limiting case where the three given points A, B, and C approach O that is the pole of the given circle. In this limit, OA, OB, and OC vanish so that α_+, β_+, and γ_+ approach -1. Similarly, when A, B, and C approach the antipodal point of O, α_+, β_+, and γ_+ approach -1. In both limits,

$$-F \approx K \approx 1 + p^2 \quad \text{and} \quad G \approx H \approx -q(1 + p^2).$$

The quadratic equation in s (third equation of (7)) now becomes

$$\left(\frac{1 + qs}{q - s}\right)^2 \approx -1$$

that has no real roots. Thus, there exist A, B, and C near O (or near the antipodal point of O) so that the inscribed triangle does not exist.

It remains to prove the following lemma.

Lemma 4.2. *Let $\mathcal{C}$ be a small circle on a sphere with pole O. Let A be a point on the sphere. Let a geodesic through A intersect $\mathcal{C}$ at points M and N. Then, we have*

$$\tan\left(\frac{1}{2}\widehat{AOM}\right)\tan\left(\frac{1}{2}\widehat{AON}\right) = \frac{\sin(OA - OM)}{\sin(OA + OM)}.$$

Proof. For conciseness, let us set $OM = r$, $OA = a$, $\widehat{AOM} = x$, and $\widehat{AON} = y$ (see Figure 7). By Lemma 2.3 in [7], in the triangle AOM, we have

$$\tan(\widehat{OAM}) = \frac{2\sin(OM)}{\tan\left(\frac{1}{2}\widehat{AOM}\right)\sin(OA + OM) + \cot\left(\frac{1}{2}\widehat{AOM}\right)\sin(OA - OM)},$$

while in the triangle AON, we have

$$\tan(\widehat{OAN}) = \frac{2\sin(ON)}{\tan\left(\frac{1}{2}\widehat{AON}\right)\sin(OA + ON) + \cot\left(\frac{1}{2}\widehat{AON}\right)\sin(OA - ON)}.$$

Since $\widehat{OAN} = \widehat{OAM}$ and $OM = ON$, this implies

$$\sin(a + r)\tan\left(\frac{x}{2}\right) + \sin(a - r)\cot\left(\frac{x}{2}\right)$$
$$= \sin(a + r)\tan\left(\frac{y}{2}\right) + \sin(a - r)\cot\left(\frac{y}{2}\right).$$

Solving for $\tan\left(\frac{x}{2}\right)$ in terms of $\tan\left(\frac{y}{2}\right)$ in this quadratic equation, we find

$$\tan\left(\frac{x}{2}\right) = \tan\left(\frac{y}{2}\right) \quad \text{or} \quad \tan\left(\frac{x}{2}\right)\tan\left(\frac{y}{2}\right) = \frac{\sin(a-r)}{\sin(a+r)}.$$

When $\tan\left(\frac{x}{2}\right) = \tan\left(\frac{y}{2}\right)$, x and y differ by an integer multiple of 2π implying that M and N coincide. Here, AM is tangent to the circle, hence $\widehat{AMO} = \frac{\pi}{2}$. Thus, we have $\cos(x) = \cot(a)\tan(r)$. This implies

$$\tan\left(\frac{x}{2}\right)\tan\left(\frac{y}{2}\right) = \tan^2\left(\frac{x}{2}\right) = \frac{1-\cos x}{1+\cos x} = \frac{\sin(a-r)}{\sin(a+r)}. \qquad \square$$

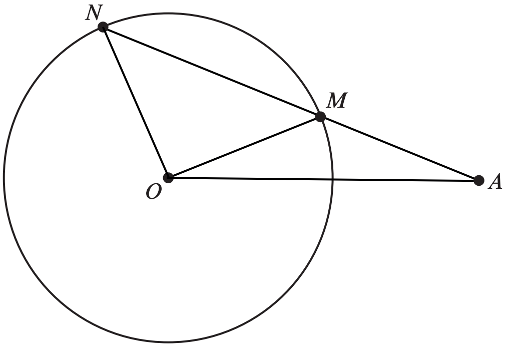


Figure 7

Acknowledgements. The author would like to thank Athanase Papadopoulos for suggesting this problem and translating the original solution due to Lagrange. The author would like to thank Vincent Alberge for critically reviewing this paper, correcting errors and suggesting significant improvements. The author acknowledges the hospitality of ICTS/Prog-ggd/2017/11 where part of this work was done.

References

[1] Pappus d'Alexandrie, *La collection mathématique,* œuvre traduite pour la premiere fois du grec en français par Paul Ver Eecke, Paris, Bruges, Desclée de Brouwer, 1933.
☞ 67

[2] J. de Castillon, “Sur une nouvelle propriété des sections coniques,” *Nouveaux mémoires de l’Académie royale des Sciences et Belles-Lettres de Berlin* 1776, 294–311.
☞ 67

[3] J.-L. Lagrange, “Solution algébrique d’un problème de géométrie,” in *Œuvres de Lagrange*, Vol. IV, Gauthier-Villars, Paris, 1868, 335–339.
☞ 67

[4] L. Euler, “Problematis cuiusdam Pappi Alexandrini constructio,” *Acta academiae scientarum imperialis Petropolitanae* 1 (1780), 91–96. Reprinted in *Opera omnia,* Series 1, Vol. XXVI, 237–242.
`http://eulerarchive.maa.org/pages/E543.html`
☞ 67

[5] N. Fuss, “Solutio problematis geometrici Pappi Alexandrini,” *Acta academiae scientarum imperialis Petropolitanae* 1 (1780), 97–104.
☞ 68

[6] E.-C. Catalan, *Théorèmes et problèmes de géométrie élémentaire,* Dunod, Paris, 1897.
☞ 68

[7] H. Senapati, “Monotonicity in spherical and hyperbolic triangles,” in *Eighteen essays in non-Euclidean geometry* (V. Alberge and A. Papadopoulos, eds.), European Mathematical Society (EMS), Zürich, 2019, Chapter 6, 81–91.
☞ 77

Chapter 6

Monotonicity in spherical and hyperbolic triangles

Himalaya Senapati

Contents

1. Introduction 81
2. Monotonicity in spherical triangles 82
3. Monotonicity in hyperbolic triangles 88

References 91

1 Introduction

The *Spherics* by Menelaus of Alexandria (1^{st}–2^{nd} century CE) is a treatise on spherical geometry in the spirit of Euclid's *Elements*. The propositions follow a logical order, starting with the fundamental properties of a spherical triangle and then exploring further and more involved properties. A critical edition of this treatise with the first English translation accompanied by mathematical and historical commentaries has recently been published [1]. We start by recalling Proposition 27 in this treatise. This is a comparison theorem between the base of a spherical triangle and the great circle arc joining the midpoints of the two legs. The theorem expresses a property of the sphere which later became known as Busemann's criterion for positive curvature in metric spaces [2].

Theorem 1.1 (*Spherics,* Proposition 27). *In any spherical triangle ABC, let D and E be the midpoints of AB and BC, and let DE be the arc joining them. Then $DE > \frac{1}{2}AC$.*

We now state Propositions 28 and 29 of the same treatise. These are comparison theorems of a similar sort, now comparing angles instead of edges, when a spherical triangle is divided in the same way by a great circle arc joining midpoints of two sides.

Theorem 1.2 (*Spherics,* Proposition 28). *Let ABC be a spherical triangle such that the angle B is at least equal to a right angle, let D and E be the midpoints of AB and BC respectively and let us join DE. Then, $\widehat{BDE} < \widehat{BAC}$ and $\widehat{BED} < \widehat{BCA}$.*

Theorem 1.3 (*Spherics,* Proposition 29). *Let ABC be a triangle such that the angle A is at least equal to a right angle, let E be the midpoint of BC, D the midpoint of AB and G the midpoint of AC. We join ED, DG and GE. Then, $\widehat{BDE} < \widehat{BAC}$ and $\widehat{CGE} < \widehat{CAB}$.*

In what follows, we first prove a monotonicity result in the lineage of Propositions 28 and 29. The statement is that in a spherical triangle ABC with $BC > AC$, if the sides CA and CB are increased while holding the angle C and the ratio CA/CB constant, the angle $\widehat{CAB}$ increases monotonically. As a corollary, we prove that the angle subtended at the base of a spherical triangle is greater than its homologue subtended at the geodesic joining the midpoints of the legs if the adjacent leg is shorter than the other leg. Here, the assumption on angles in Propositions 28 and 29 has been replaced with the assumption on sides. We then extend this comparison theorem to hyperbolic geometry proving that in a hyperbolic triangle ABC with $BC > AC$, if the sides CA and CB are increased while holding the angle C and the ratio CA/CB constant, then the angle $\widehat{CAB}$ decreases monotonically.

2 Monotonicity in spherical triangles

A spherical triangle is a triangle formed on the surface of a sphere by three great circle arcs, each less than a semi circle, intersecting pairwise at three points called vertices. The length of a side is given by the angle subtended at the centre of the sphere by the two rays connecting the centre to the endpoints of the side. The angle at a vertex is the dihedral angle between the planes containing the two sides. We begin by proving that the angles subtended at the vertices are all less than π.

Lemma 2.1. *In a spherical triangle, the angles subtended at the vertices are less than π.*

Proof. Let ABC be a spherical triangle so that AXB is less than a semicircle (Figure 1). Complete the great circle $AXBY$ where we have BYA greater than a semicircle. We prove that the angle subtended by the great circle arc AXB at C is less than the one by BYA at C. Since the two angles add up to 2π, this shows that the angle C subtended by AXB is less than π. The antipodal points of A and B (called A' and B' respectively) lie on the arc BYA since AA' and BB' are semicircles while the arc AXB is less than a semi circle. The great circle arcs AC and BC, when produced, pass through A' and B' respectively. Thus, the minor great circle arc $B'A'$ (contained in BYA) subtends the same angle at C as the arc AXB implying that the angle subtended by BYA at C is greater than the one by AXB. ☐

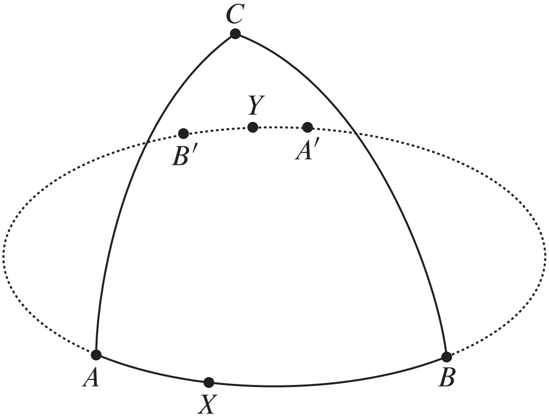


Figure 1. In a spherical triangle ABC, AB subtends an angle less than π at C

We now prove an inequality involving the sine function that will be used in proving Theorem 2.4.

Lemma 2.2. *If* $0 < a < \pi$ *and* $0 < k < 1$, *then* $k\sin(a) < \sin(ak)$.

Proof. Proving this lemma is equivalent to showing that

$$\frac{\sin(ak)}{k} - \sin(a) > 0 \quad \text{or} \quad \int_0^a (\cos(kx) - \cos(x))\, dx > 0.$$

We note that $0 < kx < x < \pi$ since $0 < a < \pi$ and $0 < k < 1$. Since cosine is monotonically decreasing in the interval $(0, \pi)$, this implies $\cos(kx) > \cos(x)$, hence proving the lemma. ☐

Below, we deduce a formula for the value of an angle at the base of a spherical triangle when the length of its two legs and the angle in between are given.

Lemma 2.3. *Let ABC be a spherical triangle where the sides AB, BC, and CA are denoted by c, a, and b respectively. Then, for $A \neq \frac{\pi}{2}$,*

$$\tan(A) = \frac{2\sin a}{\tan\left(\frac{1}{2}C\right)\sin(a+b) - \cot\left(\frac{1}{2}C\right)\sin(a-b)}$$

while for $A = \frac{\pi}{2}$,

$$\sin(a+b) \times \tan^2\left(\frac{1}{2}C\right) = \sin(a-b).$$

Proof. We first recall that when $A + B \neq \pi$ and $a + b \neq \pi$, Napier's analogies give

$$\tan\left(\frac{1}{2}(A+B)\right) = \frac{\cot\left(\frac{1}{2}C\right)\cos\left(\frac{1}{2}(a-b)\right)}{\cos\left(\frac{1}{2}(a+b)\right)}$$

and

$$\tan\left(\frac{1}{2}(A-B)\right) = \frac{\cot(\frac{1}{2}C)\sin\left(\frac{1}{2}(a-b)\right)}{\sin\left(\frac{1}{2}(a+b)\right)}.$$

Here, notice that $0 < \tan\left(\frac{1}{2}C\right) < \infty$ since $0 < \frac{1}{2}C < \frac{\pi}{2}$. We now proceed to the proof.

CASE 1: $A \neq \frac{\pi}{2}$. We first consider $A + B \neq \pi$. By the law of sines, $a + b = \pi$ would imply $A = B$ leading to $A = \frac{\pi}{2}$. Hence, $a + b \neq \pi$. We now rewrite $\tan(A)$ as $\tan\left(\frac{1}{2}(A+B) + \frac{1}{2}(A-B)\right)$ to get the relation

$$\begin{aligned}\tan(A) &= \frac{\tan\left(\frac{1}{2}(A+B)\right) + \tan\left(\frac{1}{2}(A-B)\right)}{1 - \tan\left(\frac{1}{2}(A+B)\right)\tan\left(\frac{1}{2}(A-B)\right)} \\ &= \frac{\cot\left(\frac{1}{2}C\right) \times \frac{2\sin(a)}{\sin(a+b)}}{1 - \cot^2\left(\frac{1}{2}C\right) \times \frac{\sin(a-b)}{\sin(a+b)}} \\ &= \frac{2\sin(a)}{\tan\left(\frac{1}{2}C\right)\sin(a+b) - \cot\left(\frac{1}{2}C\right)\sin(a-b)}.\end{aligned}$$

When $A + B = \pi$, the law of sines implies $a + b = \pi$ or $a = b$. However the latter is not possible since it implies $A = B$ leading to $A = \frac{\pi}{2}$. Substituting $A + B = \pi$ and $a + b = \pi$ in the second Napier's analogy, we find

$$\tan\left(A - \frac{\pi}{2}\right) = \frac{\cot\left(\frac{1}{2}C\right) \times \sin(\pi - 2a)}{2\sin(a)}$$

implying

$$\tan(A) = \frac{2\sin(a)}{\cot\left(\frac{1}{2}C\right) \times \sin(b-a)}.$$

Since $\sin(a+b) = 0$, this satisfies the relation given in the lemma.

CASE 2: $A = \frac{\pi}{2}$. When $B \neq \frac{\pi}{2}$, then $1 - \tan\left(\frac{1}{2}(A+B)\right) \times \tan\left(\frac{1}{2}(A-B)\right)$ vanishes implying

$$1 - \cot^2\left(\frac{1}{2}C\right)\frac{\cos\left(\frac{1}{2}(a-b)\right)}{\cos\left(\frac{1}{2}(a+b)\right)} \times \frac{\sin\left(\frac{1}{2}(a-b)\right)}{\sin\left(\frac{1}{2}(a+b)\right)} = 0$$

$$\iff \sin(a+b) \times \tan^2\left(\frac{1}{2}C\right) = \sin(a-b).$$

When $B = \frac{\pi}{2}$, by the law of cosines,

$$0 = \cos(A) = -\cos(B)\cos(C) + \sin(B)\sin(C)\cos(a) = \sin(C)\cos(a).$$

Hence, $\cos a = 0$ implying $a = \frac{\pi}{2}$. Since $A = B$, $b = a = \frac{\pi}{2}$. Thus, $\sin(a-b) = \sin(a+b) = 0$ so that the implied relation is satisfied. Hence, we have proved the lemma. ☐

We now state and prove a monotonicity result in spherical geometry in the lineage of Propositions 28 and 29. The idea is to show that the derivative of the angle involved with respect to a side is positive and hence the angle increases when the side increases.

Theorem 2.4. *Let ABC be a spherical triangle with $BC > AC$. Let D and E be points on AC and BC so that $CD/CA = CE/CB$. If $CD < CA$, then $\widehat{CDE} < \widehat{CAB}$.*

Proof. We refer to Figure 2. In a spherical triangle ABC, we denote the sides AB, BC, and CA by c, a, and b. Notice that the triangle is completely determined by C, a, and b, or equivalently, by C, $\frac{b}{a}$, and a. Here, $0 < a < \pi$, and by Lemma 2.1, $0 < C < \pi$. We set $\frac{b}{a} = k$ where $0 < k < 1$ since $b < a$. In the following, we first prove that A is a continuous function on the interval $0 < a < \pi$. We then show that for $A \neq \frac{\pi}{2}$, $\frac{\partial A}{\partial a} > 0$ while holding the angle C and the ratio k fixed. Next, we prove that A takes the value $\frac{\pi}{2}$ for at most one value of a for fixed C and k. This proves the theorem.

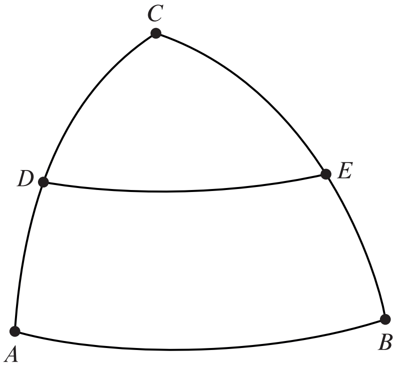


Figure 2. ABC is a spherical triangle, $BC > AC$ and $CD/CA = CE/CB$

We set

$$D = \tan\left(\frac{1}{2}C\right)\sin(a + ak) - \cot\left(\frac{1}{2}C\right)\sin(a - ak).$$

By Lemma 2.3, $A = \frac{\pi}{2}$ when $D = 0$ and since D is the restriction of a holomorphic function, it has a finite number of zeroes. When $D \neq 0$, $\tan(A) = 2\sin(a)/D$ and A is continuous in a. Thus, to show that A is continuous on the interval $0 < a < \pi$, we need to show that A is continuous at all points p in the interval where $D(p) = 0$. At such a point p, $A = \frac{\pi}{2}$. Since $2\sin a \neq 0$ for $0 < a < \pi$, $\lim_{a\to p} D(a) \to 0$ implies $\tan A \to \pm\infty$ or $A \to \frac{\pi}{2}$. Hence, A is continuous on the interval $0 < a < \pi$.

We note that

$$A = \begin{cases} \arctan(\tan(A)) & \text{for } 0 < A < \frac{\pi}{2}, \\ \pi + \arctan(\tan(A)) & \text{for } \frac{\pi}{2} < A < \pi. \end{cases}$$

This implies that for $A \neq \frac{\pi}{2}$,

$$\frac{\partial A}{\partial a} = \frac{1}{1 + \tan^2(A)} \times \frac{\partial}{\partial a}\tan(A).$$

Therefore, for $A \neq \frac{\pi}{2}$,

$$\frac{\partial}{\partial a}(\tan(A)) > 0 \implies \frac{\partial(A)}{\partial a} > 0.$$

We now show that $\frac{\partial}{\partial a}\tan(A) > 0$. Since $\tan(A) = 2\sin(a)/D$, this holds true if

$$D\frac{\partial}{\partial a}(2\sin(a)) - 2\sin(a)\frac{\partial D}{\partial a} > 0.$$

The left-hand side of the inequality is equal to $D_1 + D_2$ where

$$\begin{aligned} D_1 &= 2\tan\left(\frac{1}{2}C\right)(\sin(a+ak)\cos(a) - (1+k)\cos(a+ak)\sin(a)) \\ &= 2\tan\left(\frac{1}{2}C\right)(\sin(ak) - k\sin(a)\cos(a+ak)) \end{aligned}$$

and

$$\begin{aligned} D_2 &= 2\cot\left(\frac{1}{2}C\right)((1-k)\cos(a-ak)\sin(a) - \cos(a)\sin(a-ak)) \\ &= 2\cot\left(\frac{1}{2}C\right)(\sin(ak) - k\sin(a)\cos(a-ak)). \end{aligned}$$

Since $0 < C < \pi$, $\tan\left(\frac{1}{2}C\right)$ and $\cot\left(\frac{1}{2}C\right)$ are both strictly positive. As $0 < a < \pi$ and $0 < k < 1$, by Lemma 2.2, we have $\sin(ak) > k\sin a$ and thus

$$\sin(ak) > k\sin(a)\cos(a+ak) \quad \text{and} \quad \sin(ak) > k\sin(a)\cos(a-ak).$$

This implies $\frac{\partial}{\partial a}\tan(A) = D_1 + D_2 > 0$ since $D_1 > 0$ and $D_2 > 0$. Thus, we have proved that for $A \neq \frac{\pi}{2}$, $\frac{\partial A}{\partial a} > 0$.

Since A can take the value $\frac{\pi}{2}$ at finitely many points, the inequality $\frac{\partial A}{\partial a} > 0$ when $A \neq \frac{\pi}{2}$ implies that for fixed C and k, $A = \frac{\pi}{2}$ for at most one value of a. We have thus proved the theorem. □

The assumption that the adjacent leg is shorter than the farther one relaxes the condition that one angle at the vertices be greater than $\frac{\pi}{2}$ in Propositions 28 and 29 of Menelaus (Theorems 1.2 and 1.3). In particular, in the following corollary, $C > \frac{\pi}{2}$ has been replaced by $CB > CA$.

Corollary 2.5. *Let ABC be a spherical triangle and let D and E be the midpoints of CA and CB respectively. Let us join DE. If $CB > CA$, then $\widehat{CDE} < \widehat{CAB}$.*

Remark 2.6. In Propositions 28 and 29 of the *Spherics*, an angle at the vertex was assumed greater than $\frac{\pi}{2}$. In Theorem 2.4, while proving monotonicity of a base angle, we made the assumption that the adjacent leg is shorter than the distant one. More precisely, $CB > CA$ was needed to show monotonicity of the angle A. We now show that these conditions are crucial.

In Figure 3, $CAC'B$ is a lune. We first fix a constant $\epsilon \ll 1$. We then find points A and B on the boundary of the lune and arbitrarily close to C' with the condition $CB = (1-\epsilon)CA$. Hence, $\widehat{CAB} \approx C' = C$ and $\widehat{CBA} \approx \pi$. Let E be the midpoint of CA and F that of CB so that $\widehat{CEF} \approx \frac{\pi}{2}$. If the angle C were to be less than a right angle, it would imply

$$\widehat{CAB} \approx C < \frac{\pi}{2} \approx \widehat{CEF}$$

in violation of Propositions 28 and 29. This necessitates the condition that the corresponding angle at the vertex in a spherical triangle be greater than $\frac{\pi}{2}$.

We now find two more points X and Y arbitrarily close to C on great circle arcs CAC' and CBC' satisfying $CY = (1-\epsilon)CX$ (note that $CX/CA = CY/CB$). Here, $\widehat{CXY} \approx \widehat{CYX} \approx \frac{1}{2}(\pi - C)$. Thus, when the angle C is less than $\frac{\pi}{2}$, we have $\widehat{CXY} < \widehat{CEF}$ but $\widehat{CEF} > \widehat{CAB}$. Hence, no monotonicity result can be established in this setting. We have thus seen that the required conditions on angles in Propositions 28 and 29 of the *Spherics* and that on sides in Theorem 2.4 are crucial.

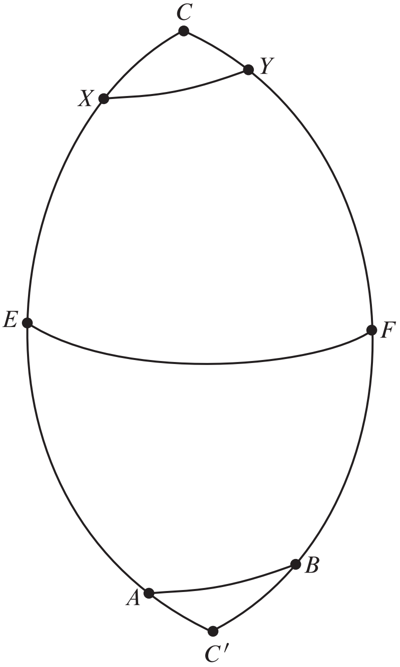


Figure 3. C and C' are antipodal points. E and F are midpoints of CA and CB and $CX/CY = CA/CB$.

3 Monotonicity in hyperbolic triangles

In this section, we obtain a theorem on monotonicity in hyperbolic triangles that is analogous to Theorem 2.4. Before proceeding, we first prove a relation among angles and sides in a hyperbolic triangle that will be used in proving the theorem.

Lemma 3.1. *Let ABC be a hyperbolic triangle. Let the sides AB, BC, and CA be denoted by c, a, and b. Then for $A \neq \frac{\pi}{2}$,*

$$\tan(A) = \frac{\sin(C)}{\sinh(b)\coth(a) - \cosh(b)\cos(C)}$$

while for $A = \frac{\pi}{2}$, $\sinh(b)\coth(a) = \cosh(b)\cos(C)$.

Proof. By the law of cosines, we have

$$\cosh(a) = \cosh(b)\cosh(c) - \sinh(b)\sinh(c)\cos(A).$$

In this equation, we substitute

$$\cosh(c) = \cosh(a)\cosh(b) - \sinh(a)\sinh(b)\cos(C)$$

and

$$\sinh(c) = \frac{\sinh(a)}{\sin(A)} \times \sin(C)$$

to get

$$\begin{aligned}&\sinh(a)\sinh(b)\cot(A)\sin(C)\\ &\quad = \cosh(a)\sinh^2(b) - \sinh(a)\sinh(b)\cosh(b)\cos(C).\end{aligned}$$

Dividing both sides by $\sinh(a)\sinh(b)$, we obtain

$$\cot(A)\sin(C) = \coth(a)\sinh(b) - \cosh(b)\cos(C).$$

This implies that for $A = \frac{\pi}{2}$, $\coth(a)\sinh(b) = \cosh(b)\cos(C)$ while for $A \neq \frac{\pi}{2}$,

$$\tan(A) = \frac{\sin(C)}{\sinh(b)\coth(a) - \cosh(b)\cos(C)}. \qquad \square$$

.

We now state and prove the theorem on monotonicity in hyperbolic triangles.

Theorem 3.2. *Let ABC be a hyperbolic triangle with $BC > AC$. Let D and E be points on AC and BC so that $CD/CA = CE/CB$. Furthermore, let $CD < CA$. Then, $\widehat{CDE} > \widehat{CAB}$.*

Proof. In Figure 4, we set $BC = a$ and $CA = b$. Proceeding similarly to the spherical case, it follows from Lemma 3.1 that A is continuous in a. We first show that for $A \neq \frac{\pi}{2}$, $\frac{\partial A}{\partial a} < 0$ while holding the angle C and the ratio $\frac{b}{a}$ fixed. We then show that there is at most one value of a for fixed C and $\frac{b}{a}$ for which $A = \frac{\pi}{2}$. This, along with continuity of A in a proves the theorem.

By Lemma 3.1, for $A \neq \frac{\pi}{2}$, we have

$$\tan(A) = \frac{\sin(C)}{\sinh(b)\coth(a) - \cosh(b)\cos(C)}.$$

Using quotient rule and setting $\frac{b}{a} = k =$ constant,

$$\frac{\partial}{\partial a}(\tan(A)) = -\frac{k\cosh(ak)\coth(a) - (k\cos(C) + \operatorname{csch}^2(a))\sinh(ak)}{(\sinh(b)\coth(a) - \cosh(b)\cos(C))^2}\sin(C).$$

Since $0 < C < \pi$, $\sin C > 0$. Thus, $\frac{\partial}{\partial a}\tan(A) < 0$ if

$$k\cosh(ak)\coth(a) - (k\cos(C) + \operatorname{csch}^2(a))\sinh(ak) > 0.$$

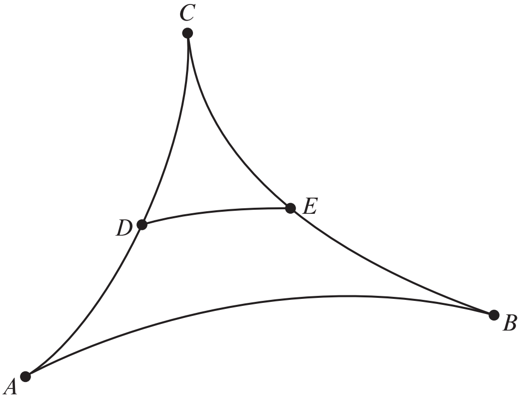


Figure 4

As the left-hand side is linear in $\cos(C)$, we need to show that

$$\Delta_\pm = k\cosh(ak)\coth(a) - (\pm k + \operatorname{csch}^2(a))\sinh(ak) > 0.$$

However, $\Delta_+ > 0$ implies $\Delta_- > \Delta_+ > 0$ since $\sinh(ak) > 0$. To prove $\Delta_+ > 0$ is equivalent to showing $\Delta = (\sinh^2(a))\Delta_+ > 0$. We find

$$\frac{\partial\Delta}{\partial a} = (2-k)k\sinh(a)\sinh(a - ak)$$

so that the derivative never vanishes when $0 < k < 1$ and $0 < a < \infty$. This implies that Δ does not have a local extremum in this region. At $k = 0$ or 1, $\Delta = 0$ while

$$\lim_{a\to 0}\Delta = 0 \quad \text{and} \quad \lim_{a\to\infty}\Delta = \infty.$$

This implies that Δ does not change sign in the region $0 < k < 1$, $0 < a < \infty$ and hence $\Delta > 0$. We have now shown that $\frac{\partial A}{\partial a} < 0$ for $A \neq \frac{\pi}{2}$.

We now show that for fixed C and $\frac{b}{a}$, there is at most one value of a for which $A = \frac{\pi}{2}$ and that A is a continuous function of a. We note that, by Lemma 3.1, $A = \frac{\pi}{2}$ when

$$\sinh(b)\coth(a) - \cosh(b)\cos(C) = 0 \iff \tanh(a) \times \tanh(ak) = \cos(C).$$

Since both $\tanh a$ and $\tanh ak$ are positive and monotonically increasing functions for $a > 0$ and $0 < k < 1$, their product is also an increasing function. Therefore, $\tanh(a) \times \tanh(ak) = \cos(C)$ has at most one solution for fixed values of k and C. We have thus shown that $\frac{\partial A}{\partial a} < 0$ for $A \neq \frac{\pi}{2}$ and that $A = \frac{\pi}{2}$ for at most one value of a for fixed C and $\frac{b}{a}$. Hence, we have proved the theorem. □

Acknowledgements. The author would like to thank Athanase Papadopoulos and Vincent Alberge for critically reading through this paper, correcting errors and suggesting significant improvements.

References

[1] R. Rashed and A. Papadopoulos, *Menelaus' spherics*: *early translation and al-Māhānī/al-Harawī's version,* Critical edition with English translation and historical and mathematical commentaries, Series Scientia Graeco-Arabica 21, de Gruyter, Berlin, 2017.
Zbl 06908518 ☞ 81

[2] H. Busemann, "Spaces with non-positive curvature," *Acta Math.* 80 (1948), 259–310.
MR 0029531 Zbl 0038.10005 ☞ 81

Chapter 7

De Tilly's mechanical view on hyperbolic and spherical geometries

Dmitriy Slutskiy

Contents

1. Introduction 93
2. Functions circ and eq: modern calculation 95
3. The functions circ and eq: ancient approach 96
4. Corollaries 107

References 110

1 Introduction

The fifth Euclidean postulate has been a subject of controversy between mathematicians regarding its necessity and independence from the other axioms since Euclid stated it over two thousand years and until the discovery of non-Euclidean geometry in the 19th century. Many famous geometers and philosophers tried to show that this axiom is a consequence of the other axioms, without success. In 1733 Girolamo Saccheri [4],, the Italian priest and mathematician who taught theology and mathematics at the University of Pavia, considered the first four Euclidean postulates together with the negation of the fifth, and tried to show the inconsistency of this system of postulates, and thus to prove the necessity of the fifth postulate, that is, to prove that it logically follows from the other axioms. Saccheri thought he succeeded to realize his idea but he made a mistake in his proof. In 1766, Johann Heinrich Lambert published a similar investigation, but he did not find any contradiction in the new system of postulates and did not emit any definite conclusion; see [2] and [3] for the details.

The fifth postulate is equivalent to the fact that the sum of the angles in a triangle is equal to π. Lambert knew that the angle sum in a spherical triangle is greater than π, and that the area of a spherical triangle is, up to a constant factor, equal to the angle sum minus π. In his 1766 memoir, he made an extensive study of quadrilaterals with three right angles in the three geometries (the Euclidean, the spherical and the hyperbolic), which are called today *Lambert quadrilaterals*, or *Ibn al-Haytham–Lambert* quadrilaterals. He showed that the angle sum in a triangle in the (hypothetical) hyperbolic plane must be less than π, and that the area of such a triangle has to be proportional to the angle defect, that is, π minus the angle sum, a situation analogous to the one in spherical geometry but where inequalities have to be reversed: angle sum is greater than π and area is proportional to angle excess, that is, angle sum minus π.

Joseph-Marie de Tilly (1837–1906), a Belgian artillery colonel and professor at the Royal Military School in Brussels worked extensively on non-Euclidean geometry. After János Bolyai, Carl Friedrich Gauss, and Nikolai Ivanovich Lobachevsky and independently of them, he rediscovered hyperbolic geometry, in the 1860s. Like his famous predecessors, de Tilly found a geometry where the Euclidean fifth postulate is replaced by its negation, but unlike them, he used reasonings from mechanics in his investigation of this new geometry which incidentally he called "abstract geometry"[1] in order to distinguish it from from the "real" Euclidean geometry and from Lobachevsky's "imaginary geometry," although the Lobachevsky geometry and his geometry amount to the same.

The key idea of de Tilly was to use a velocity-addition formula which was known in kinematics, assuming that at every point of the two-dimensional space under consideration, the tangent space is the vector space $\mathbb{R}^2$ with the standard Euclidean norm.[2] In his 1870 paper titled "Études de mécanique abstraite" [1], he applies his kinematical approach to the study of non-Euclidean geometry. In the first and second parts of his paper, de Tilly applies his method to find several formulae relating the length of a curve at distance r from a geodesic (a function he calls $\mathrm{eq}(r)$) and of the length of a circle of radius r (a function he calls $\mathrm{circ}(r)$) on his "abstract" 2-plane. Combining the obtained relations, he gives an original proof of formulae (3) below in hyperbolic geometry, and he also deduces the Laws of Sines and Cosines in spherical space. It is important to note that by the time he was working on this subject, the Poincaré model of the hyperbolic plane was not yet known, and de Tilly could not work in models. He used other methods we shall explain below. The third part of his paper is mostly dedicated to the basics of mechanics in the new "abstract geometry." For instance, de Tilly studies compositions of translations and rotations and establishes the formulae for the transition from one inertial reference frame to another. A quick overview of de Tilly's paper is given in [5, pp. 85–110].

---

[1] "Géométrie abstraite" in French.

[2] We are using here modern mathematical terminology.

It is may be useful to recall that the terminology and notions universally known and accepted by mathematicians has changed since the 19th century. Therefore not all proofs in de Tilly's paper [1] are clear today, although there is no doubt that this paper was totally understandable by the mathematicians 150 years ago. In Section 2, we give formulae for the functions $\mathrm{eq}(r)$ and $\mathrm{circ}(r)$ in the Euclidean, hyperbolic, and spherical cases. Alternatively, it is not difficult to obtain them by direct calculations with the use of modern techniques. In Section 3, we give several examples of de Tilly's reasoning to explain the main principles. Quite standard manipulations combined with the results of Sections 2 and 3 imply a unified proof of the Laws of Sines and Cosines in both hyperbolic and spherical spaces. This is the content of Section 4.

2 Functions circ and eq: modern calculation

Following de Tilly's paper [1], let us introduce, in the Euclidean, hyperbolic, or spherical plane, the functions

- $\mathrm{circ}(r)$ equal to the length of a circle of radius $r \geq 0$, defined whenever the corresponding circle exists;
- $\mathrm{eq}(r)$ equal to the length of a curve at distance $r \geq 0$ from a geodesic segment of length 1, again defined whenever the corresponding equidistant curve exists.

In what follows, when we talk about a "space," it is understood that it is 2-dimensional.

For the Euclidean space, let us denote the corresponding functions by $\mathrm{circ}_{\mathbb{R}}(r)$ and $\mathrm{eq}_{\mathbb{R}}(r)$, for the spherical space by $\mathrm{circ}_{\mathbb{S}}(r)$ and $\mathrm{eq}_{\mathbb{S}}(r)$, and for the hyperbolic space by $\mathrm{circ}_{\mathbb{H}}(r)$ and $\mathrm{eq}_{\mathbb{H}}(r)$.

It is elementary that in Euclidean space we have the formulae

$$\mathrm{circ}_{\mathbb{R}}(r) = 2\pi r \quad \text{and} \quad \mathrm{eq}_{\mathbb{R}}(r) = 1 \tag{1}$$

for all $r \geq 0$.

A direct calculation on a sphere of curvature $K = M^2$, $M > 0$, gives

$$\mathrm{circ}_{\mathbb{S}}(r) = \frac{2\pi}{M}\sin(Mr) \quad \text{and} \quad \mathrm{eq}_{\mathbb{S}}(r) = \cos(Mr) \tag{2}$$

defined for $0 \leq r \leq \frac{\pi}{M}$ and $0 \leq r \leq \frac{\pi}{2M}$, respectively.

A similar calculation in the Poincaré model of the hyperbolic plane of curvature $K=-M^2$, $M>0$, shows that

$$\mathrm{circ}_{\mathbb{H}}(r)=\frac{2\pi}{M}\sinh(Mr) \quad \text{and} \quad \mathrm{eq}_{\mathbb{H}}(r)=\cosh(Mr) \tag{3}$$

defined for all $r\geq 0$.

Thanks to (1)–(3) we can naturally extend the functions eq and circ to the whole real line $\mathbb{R}$. Unifying (1)–(3) we conclude that for a space of constant curvature $K\in\mathbb{R}$ the following formulae hold:

$$\mathrm{circ}_K(r)=\frac{2\pi}{\sqrt{K}}\sin(\sqrt{K}r) \quad \text{and} \quad \mathrm{eq}_K(r)=\cos(\sqrt{K}r), \tag{4}$$

$r\in\mathbb{R}$ (for $K<0$ we consider the branch $\sqrt{K}=i\sqrt{|K|}$).

By formulae (4), in a space of arbitrary constant curvature K, $K\in\mathbb{R}$, the "extended" functions eq and circ satisfy the following properties:

$$\mathrm{circ}(r)=-\mathrm{circ}(-r) \quad \text{and} \quad \mathrm{eq}(r)=\mathrm{eq}(-r), \quad \text{for all } r\in\mathbb{R}. \tag{5}$$

In his work, de Tilly supposed that, in contrast to the Euclidean case, the space he considered verifies either $\mathrm{eq}^2(r)\geq 1$ or $0\leq \mathrm{eq}^2(r)\leq 1$. These inequalities easily follow from (2) and (3). Like de Tilly, in what follows, we shall deal exclusively with his hypotheses on $\mathrm{circ}(r)$ and $\mathrm{eq}(r)$ and we shall not use the explicit expressions (2) and (3).

3 The functions circ and eq: ancient approach

In this section, we give several of de Tilly's "mechanical" ideas applied to geometry. This will allow us to obtain formulae relating the functions $\mathrm{eq}(r)$ and $\mathrm{circ}(r)$ commonly in the Euclidean, hyperbolic, and spherical spaces without the use of any model for these spaces.

Let us recall two main kinematic principles used by de Tilly.

Principle 1. If a point P moves along a geodesic l with a constant speed v then the absolute value v' of the velocity of a point P' moving along an equidistant curve l' at distance r from l such that the geodesic line passing through P and P' remains orthogonal to l at any time satisfies the formula

$$v'=v\,\mathrm{eq}(r).$$

Note that by the definition of the function $\mathrm{eq}(r)$, if P passes any distance d along l, the point P' should pass the distance $d\,\mathrm{eq}(r)$ along l'. The velocity of P' is tangent to l' and points in the direction of the movement of P'.

Principle 2. Let us consider the rotation of a point P at a constant speed ω_P around a circle of radius 1 centered at a point O. Define the angular velocity ω around O as follows: the absolute value of ω is equal to ω_P; ω is positive if P rotates in a clockwise direction, and is negative otherwise (in contrast to the standard definition of the angular velocity). Let us now consider the rotation of a point P' at distance r around O with the angular velocity ω. By definition, a full turn around O is performed at the time $\frac{\operatorname{circ}(1)}{\omega}$. As the length of a circle of radius r is $\operatorname{circ}(r)$, we conclude that the speed of P' is $\omega\frac{\operatorname{circ}(r)}{\operatorname{circ}(1)}$. The velocity vector of P' is tangent to the circle of rotation of P' and it points in the direction of rotation of P'.

Note that both principles deal only with the first four Euclidean axioms and do not use the parallel postulate.

3.1 Moving along equidistant lines

Let l_O and l_{AB} be two perpendicular straight lines intersecting each other at O, and let A and B be two points on l_{AB} such that A lies between O and B. Denote by a the length of the segment OA, and by b the length of the segment AB.

Let us move the point O together with the line l_{AB} along the geodesic l_O in such a way that l_{AB} remains orthogonal to l_O during the translation. Thus the points A and B move along the curves c_A and c_B equidistant from l_O at distances a and $a+b$, respectively. The velocity vector $\vec{v}$ of O at any time t_0 is tangent to l_O. Denote by v the length of $\vec{v}$. Since we consider a translation along l_O, the angular velocity around O is zero. Also, the velocity vectors $\vec{v}_A$ and $\vec{v}_B$ of the points A and B at the time t_0 are tangent to the curves c_A and c_B, respectively. Denote by v_A the length of $\vec{v}_A$, by v_B the length of $\vec{v}_B$, by ω_A the angular velocity around A that (together with $\vec{v}_A$) provides the moving of A along c_A as O moves along l_O, by ω_B the angular velocity around B that (together with $\vec{v}_B$) provides the moving of B along c_B as O moves along l_O (see Figure 1).

Let us consider the mutual movement of the points O and A. By Principle 1, the speed of the point A induced by the translation of the point O is

$$v_A = v\,\operatorname{eq}(a). \tag{6}$$

On the other hand, the movement of O induced by the movement of the point A can be decomposed into the rotational and the translational parts. If the point A only translates along a geodesic line with a tangent vector $\vec{v}_A$ at A, then, as previously, in order to keep the line l_{AB} orthogonal to such a geodesic, the point O should move

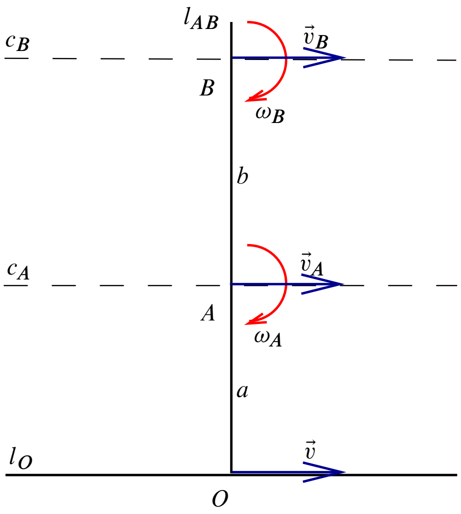


Figure 1. Moving along equidistant lines

orthogonally to l_{AB} with speed $v_A\,\mathrm{eq}(a)$ in the same direction as the point A does. We will call it a translational component of the motion of O. When A is fixed and l_{AB} rotates around A with an angular velocity ω_A then, by Principle 2, O moves along a circle centered at A of radius a with velocity vector of length $\omega_A \frac{\mathrm{circ}(a)}{\mathrm{circ}(1)}$ tangent to the considered circle, i.e orthogonal to l_{AB}. We will call it a rotational component of the motion of O. Since an angular velocity is positive if the rotation is clockwise and since the point O is located “under” A on Figure 1 (we can always turn the plane containing the lines l_{AB} and l_O so that the point O becomes “under” A), the rotational component of the motion of O has an opposite direction to the vector $\vec{v}_A$. Thus, the rotational and the translational components of the motion of O have opposite directions. Therefore, according to the kinematic velocity addition rule, we can write

$$v = v_A\,\mathrm{eq}(a) - \omega_A \frac{\mathrm{circ}(a)}{\mathrm{circ}(1)}. \tag{7}$$

Similarly, the movement of B viewed from the point O is described by

$$v_B = v\,\mathrm{eq}(a+b), \tag{8}$$

the movement of O viewed from the point B can be written as

$$v = v_B\,\mathrm{eq}(a+b) - \omega_B \frac{\mathrm{circ}(a+b)}{\mathrm{circ}(1)}, \tag{9}$$

the movement of B viewed from the point A is described by

$$v_B = v_A \operatorname{eq}(b) + \omega_A \frac{\operatorname{circ}(b)}{\operatorname{circ}(1)}, \tag{10}$$

and the movement of A viewed from the point B can be written as

$$v_A = v_B \operatorname{eq}(b) - \omega_B \frac{\operatorname{circ}(b)}{\operatorname{circ}(1)}. \tag{11}$$

Substituting (6) into (7), we get

$$\frac{\omega_A}{\operatorname{circ}(1)} = \frac{v}{\operatorname{circ}(a)} (\operatorname{eq}^2(a) - 1). \tag{12}$$

Substituting (8) into (9), we get

$$\frac{\omega_B}{\operatorname{circ}(1)} = \frac{v}{\operatorname{circ}(a+b)} (\operatorname{eq}^2(a+b) - 1). \tag{13}$$

Substituting (6), (8), and (13) into (11), we obtain

$$\operatorname{eq}(a) = \operatorname{eq}(b) \operatorname{eq}(a+b) - \frac{\operatorname{circ}(b)}{\operatorname{circ}(a+b)} (\operatorname{eq}^2(a+b) - 1).$$

Although it is not used in what follows, notice that if we set $\varphi = a + b$, $\psi = b$, $\varphi - \psi = a$, then the previous formula becomes

$$\operatorname{eq}(\varphi - \psi) = \operatorname{eq}(\psi) \operatorname{eq}(\varphi) - \frac{\operatorname{circ}(\psi)}{\operatorname{circ}(\varphi)} (\operatorname{eq}^2(\varphi) - 1). \tag{14}$$

Next, substituting (6), (8), and (12) into (10), we obtain

$$\operatorname{eq}(a+b) = \operatorname{eq}(a) \operatorname{eq}(b) + \frac{\operatorname{circ}(b)}{\operatorname{circ}(a)} (\operatorname{eq}^2(a) - 1). \tag{15}$$

By symmetry, if we change the roles of a and b in equation (15),

$$\operatorname{eq}(a+b) = \operatorname{eq}(a) \operatorname{eq}(b) + \frac{\operatorname{circ}(a)}{\operatorname{circ}(b)} (\operatorname{eq}^2(b) - 1). \tag{16}$$

Expressing $\frac{\operatorname{circ}(a)}{\operatorname{circ}(b)}$ from (15) and (16), we obtain

$$(\operatorname{eq}(a+b) - \operatorname{eq}(a) \operatorname{eq}(b))^2 = (\operatorname{eq}^2(a) - 1)(\operatorname{eq}^2(b) - 1). \tag{17}$$

Remark that for every $r > 0$ such that there exists a circle of radius r in the given space, the function $\operatorname{circ}(r)$ is positive, by definition. Looking at (15), we notice that expression $\operatorname{eq}(a+b) - \operatorname{eq}(a) \operatorname{eq}(b)$ is not positive in the spherical case (since $\operatorname{eq}^2(a) \leq 1$ in this case, see (2)), and it is not negative in the hyperbolic case (since $\operatorname{eq}^2(a) \geq 1$ in this case, see (3)). Therefore, the identity (17) implies

(s) in the spherical case

$$\mathrm{eq}(a+b) = \mathrm{eq}(a)\,\mathrm{eq}(b) - \sqrt{(\mathrm{eq}^2(a)-1)(\mathrm{eq}^2(b)-1)}, \tag{18}$$

(h) in the hyperbolic case

$$\mathrm{eq}(a+b) = \mathrm{eq}(a)\,\mathrm{eq}(b) + \sqrt{(\mathrm{eq}^2(a)-1)(\mathrm{eq}^2(b)-1)} \tag{19}$$

for all a and b where the functions circ and eq are defined. Using (2) and (3), we conclude that formulae (18) and (19) hold for all $a, b \in \mathbb{R}$.

Corollary 3.1. *Taking into account formulae* (2) *and* (3)*, expressions* (14)*, and also* (15)*,* (16)*, and* (18) *considered in spherical space generalize the formulae for* $\cos(a-b)$ *and* $\cos(a+b)$*. Moreover, expressions* (14)*, and also* (15)*,* (16)*, and* (19) *considered in hyperbolic space generalize the formulae for* $\cosh(a-b)$ *and* $\cosh(a+b)$.

3.2 Simultaneous rotation of two circles around their common center

Let l_{AB} be a straight line, and let O, A, and B be three points on l_{AB} such that A lies between O and B. Denote by a the length of the segment OA, and by b the length of the segment AB.

Let us now fix the point O (therefore, the velocity vector $\vec{v}$ of O at any time t_0 is zero) and let us rotate l_{AB} around O with the angular velocity ω. Thus the points A and B move along the circles c_A and c_B centered at O of radii a and $a+b$, respectively. By Principle 2, the velocity vectors $\vec{v}_A$ and $\vec{v}_B$ of the points A and B at any time t_0 are tangent to the circles c_A and c_B, respectively. Denote by v_A the length of $\vec{v}_A$, by v_B the length of $\vec{v}_B$, by ω_A the angular velocity around A that (together with $\vec{v}_A$) provides the rotation of A around O along c_A, by ω_B the angular velocity around B that (together with $\vec{v}_B$) provides the rotation of B around O along c_B (see Figure 2).

Reasoning in a similar way as we did in order to obtain equations (7)–(11), we get that the movement of the point A viewed from the point O is described by equation

$$v_A = \omega \frac{\mathrm{circ}(a)}{\mathrm{circ}(1)}. \tag{20}$$

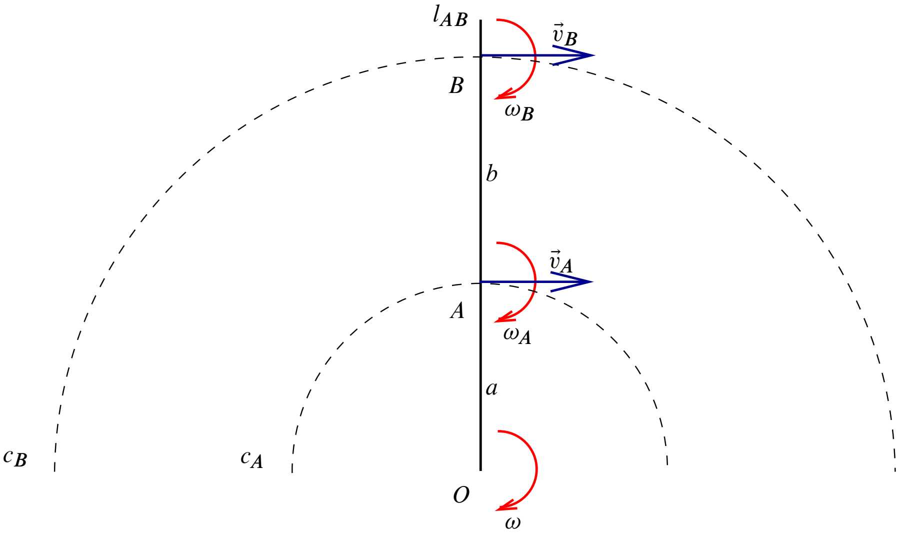


Figure 2. Simultaneous rotation of two circles around their common center

Similarly, the movement of O viewed from the point A can be written as

$$v_A \operatorname{eq}(a) - \omega_A \frac{\operatorname{circ}(a)}{\operatorname{circ}(1)} = 0, \tag{21}$$

the movement of B viewed from the point O is described by

$$v_B = \omega \frac{\operatorname{circ}(a+b)}{\operatorname{circ}(1)}, \tag{22}$$

the movement of O viewed from the point B can be written as

$$v_B \operatorname{eq}(a+b) - \omega_B \frac{\operatorname{circ}(a+b)}{\operatorname{circ}(1)} = 0, \tag{23}$$

the movement of B viewed from the point A is described by

$$v_B = v_A \operatorname{eq}(b) + \omega_A \frac{\operatorname{circ}(b)}{\operatorname{circ}(1)}, \tag{24}$$

and the movement of A viewed from the point B can be written as

$$v_A = v_B \operatorname{eq}(b) - \omega_B \frac{\operatorname{circ}(b)}{\operatorname{circ}(1)}. \tag{25}$$

Substituting (20) into (21), we get

$$\omega_A = \omega \operatorname{eq}(a). \tag{26}$$

Substituting (22) into (23), we get

$$\omega_B = \omega \operatorname{eq}(a+b). \tag{27}$$

Substituting (20), (22), and (27) into (25), we obtain

$$\operatorname{circ}(a) = \operatorname{eq}(b)\operatorname{circ}(a+b) - \operatorname{circ}(b)\operatorname{eq}(a+b).$$

Although it is not used in what follows, notice that if we set $\varphi = a+b$, $\psi = b$, $\varphi - \psi = a$, then the previous formula can be rewritten as follows:

$$\operatorname{circ}(\varphi - \psi) = \operatorname{circ}(\varphi)\operatorname{eq}(\psi) - \operatorname{eq}(\varphi)\operatorname{circ}(\psi). \tag{28}$$

Substituting (20), (22), and (26) into (24), we obtain

$$\operatorname{circ}(a+b) = \operatorname{circ}(a)\operatorname{eq}(b) + \operatorname{eq}(a)\operatorname{circ}(b). \tag{29}$$

Corollary 3.2. *Taking into account formulae* (2) *and* (3)*, expressions* (28)*, and also* (29) *considered in spherical space generalize the formulae for* $\sin(a-b)$ *and* $\sin(a+b)$*. Moreover, expressions* (28)*, and also* (29) *considered in hyperbolic space generalize the formulae for* $\sinh(a-b)$ *and* $\sinh(a+b)$*.*

3.3 Kinematics of a right-angled triangle

Let us consider a right-angled triangle $\triangle ABC$ in a plane of constant curvature $K \in \mathbb{R}$ with angles $\angle A = \frac{\pi}{2}$, $\angle B$, $\angle C$ at vertices A, B, and C, and with lengths a, b, and c of the edges BC, AC, and AB, respectively.

Lemma 3.3. *Given a triangle* $\triangle ABC$ *as above, then*

$$\operatorname{circ}(b) = \operatorname{circ}(a)\sin(\angle B) \quad \text{and} \quad \operatorname{circ}(c) = \operatorname{circ}(a)\sin(\angle C). \tag{30}$$

Proof. Let us consider a rotation of $\triangle ABC$ with angular velocity ω around the vertex C (see Figure 3).

By Principle 2, the velocity of the vertex A is a vector of length $\omega \frac{\operatorname{circ}(b)}{\operatorname{circ}(1)}$ orthogonal to CA and the velocity of the vertex B is a vector of length $\omega \frac{\operatorname{circ}(a)}{\operatorname{circ}(1)}$ orthogonal to CB.

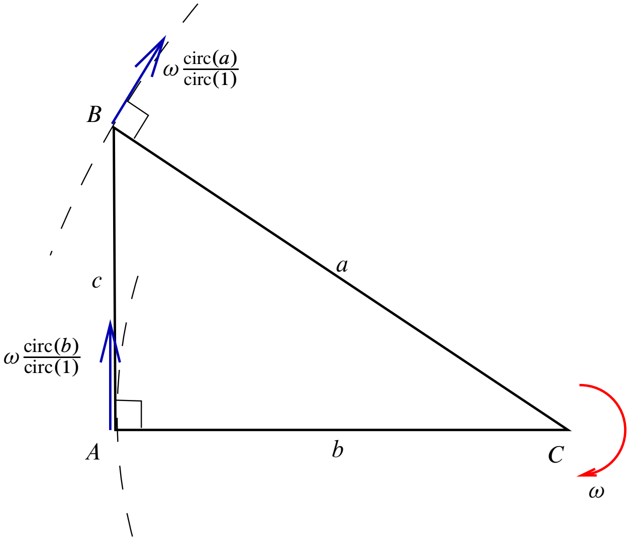

$\omega\frac{\mathrm{circ}(a)}{\mathrm{circ}(1)}$
B
a
c
$\omega\frac{\mathrm{circ}(b)}{\mathrm{circ}(1)}$
A
b
C
ω

Figure 3. Rotation of a right-angled triangle

Since the length c of the edge AB does not change during the rotation, the projections of the velocity vectors of A and B on the line AB are equal, i.e.

$$\omega\frac{\mathrm{circ}(b)}{\mathrm{circ}(1)} = \omega\frac{\mathrm{circ}(a)}{\mathrm{circ}(1)}\sin(\angle B),$$

which is equivalent to the first formula from (30).

The same reasoning applied to $\triangle ABC$ turning around the vertex B leads to the second formula from (30). □

Lemma 3.4. *Given a triangle $\triangle ABC$ as above, then*

$$\mathrm{eq}(b) = \frac{\cos(\angle B)}{\sin(\angle C)} \quad \textit{and} \quad \mathrm{eq}(c) = \frac{\cos(\angle C)}{\sin(\angle B)}. \tag{31}$$

Proof. Let us slide $\triangle ABC$ along the line A with speed v (see Figure 4).

By Principle 1, the velocity of the vertex B is a vector of length $v\,\mathrm{eq}(c)$ orthogonal to AB.

Since the length a of the edge BC does not change during the movement of $\triangle ABC$, the projections of the velocity vectors of B and C on the line BC are equal, i.e.

$$v\cos(\angle C) = v\,\mathrm{eq}(c)\cos\left(\frac{\pi}{2} - \angle B\right) \ (= v\,\mathrm{eq}(c)\sin(\angle B)),$$

which is equivalent to the second formula from (31).

The same reasoning applied to $\triangle ABC$ sliding along the line AB leads to the first formula from (31). □

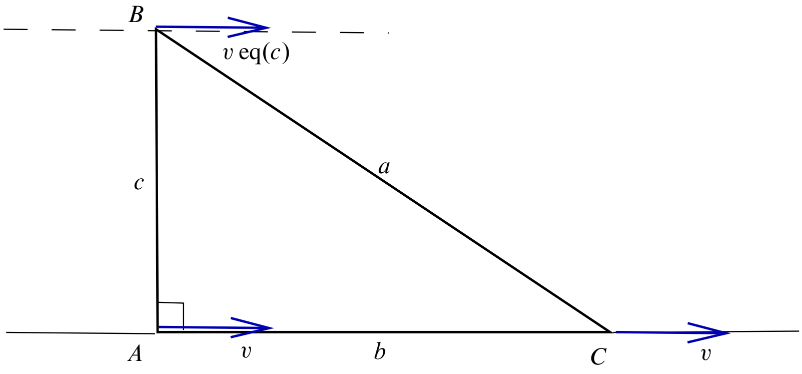


Figure 4. Translation of a right-angled triangle

Lemma 3.5. *Given a triangle $\triangle ABC$ as above, we have*

$$\operatorname{eq}(a) = \cot(\angle B)\cot(\angle C) = \frac{\cos(\angle B)}{\sin(\angle C)}\frac{\cos(\angle C)}{\sin(\angle B)} = \operatorname{eq}(b)\operatorname{eq}(c). \tag{32}$$

Proof. Let us draw a segment AD, $D \in BC$, such that AD is orthogonal to BC, and let $\alpha{=}\angle BAD$.

Applying Lemma 3.4 to the triangles $\triangle ABD$ and $\triangle ACD$, we get:

$$\operatorname{eq}(AD) = \frac{\cos(\angle B)}{\sin(\alpha)} = \frac{\cos(\angle C)}{\cos(\alpha)}.$$

Hence,

$$\tan\alpha = \frac{\cos(\angle B)}{\cos(\angle C)}.$$

By construction, $\alpha < \frac{\pi}{2}$, $\angle B < \frac{\pi}{2}$, and $\angle C < \frac{\pi}{2}$, and we can apply expressions

$$\sin(\alpha) = \frac{\tan(\alpha)}{\sqrt{1+\tan^2(\alpha)}} \quad \text{and} \quad \cos(\alpha) = \frac{1}{\sqrt{1+\tan^2(\alpha)}}$$

in order to obtain formulae

$$\cos(\alpha) = \frac{\cos\angle C}{\sqrt{\cos^2(\angle B)+\cos^2(\angle C)}}, \quad \sin(\alpha) = \frac{\cos(\angle B)}{\sqrt{\cos^2(\angle B)+\cos^2(\angle C)}}.$$

Therefore, by Lemma 3.4 applied again to the triangles $\triangle ABD$ and $\triangle ACD$, we have

$$\mathrm{eq}(BD) = \frac{\cos(\alpha)}{\sin(\angle B)} = \frac{\cos(\angle C)}{\sin(\angle B)\sqrt{\cos^2(\angle B) + \cos^2(\angle C)}}. \tag{33}$$

and

$$\mathrm{eq}(DC) = \frac{\sin(\alpha)}{\sin(\angle C)} = \frac{\cos(\angle B)}{\sin(\angle C)\sqrt{\cos^2(\angle B) + \cos^2(\angle C)}}. \tag{34}$$

Recall that, from expressions (18) and (19) it follows that

(s) in the spherical case

$$\mathrm{eq}(a) - \mathrm{eq}(BD)\,\mathrm{eq}(DC) = -\sqrt{(\mathrm{eq}^2(BD) - 1)(\mathrm{eq}^2(DC) - 1)}, \tag{35}$$

(h) in the hyperbolic case

$$\mathrm{eq}(a) - \mathrm{eq}(BD)\,\mathrm{eq}(DC) = \sqrt{(\mathrm{eq}^2(BD) - 1)(\mathrm{eq}^2(DC) - 1)}. \tag{36}$$

By formulae (33) and (34),

$$\mathrm{eq}(a) - \mathrm{eq}(BD)\,\mathrm{eq}(DC) = \mathrm{eq}(a) - \frac{\cos(\angle C)\cos(\angle B)}{\sin(\angle C)\sin(\angle B)(\cos^2(\angle B) + \cos^2(\angle C))}, \tag{37}$$

and, also,

$$\begin{aligned} &\sqrt{(\mathrm{eq}^2(BD) - 1)(\mathrm{eq}^2(DC) - 1)} \\ &\quad = \frac{\cos(\angle C)\cos(\angle B)\sqrt{(\cos^2(\angle B) + \cos^2(\angle C) - 1)^2}}{\sin(\angle C)\sin(\angle B)(\cos^2(\angle B) + \cos^2(\angle C))}. \end{aligned} \tag{38}$$

The angle sum in any triangle is greater than 2π in the spherical case, and is less than 2π in the hyperbolic case. Hence, for the right-angled triangle $\triangle ABC$ with $\angle A = \frac{\pi}{2}$ we get $\angle C > \frac{\pi}{2} - \angle B$ in spherical space, and $\angle C < \frac{\pi}{2} - \angle B$ in hyperbolic space. Thus,

(s) in the spherical case

$$\cos^2(\angle B) + \cos^2(\angle C) - 1 < 0,$$

(h) in the hyperbolic case

$$\cos^2(\angle B) + \cos^2(\angle C) - 1 > 0.$$

Therefore, formula (38) gives

(s) in the spherical case

$$\begin{aligned} &-\sqrt{(\mathrm{eq}^2(BD)-1)(\mathrm{eq}^2(DC)-1)} \\ &\quad = \frac{\cos(\angle C)\cos(\angle B)(\cos^2(\angle)B+\cos^2(\angle C)-1)}{\sin(\angle C)\sin(\angle B)(\cos^2(\angle B)+\cos^2(\angle C))}, \end{aligned} \tag{39}$$

(h) in the hyperbolic case

$$\begin{aligned} &\sqrt{(\mathrm{eq}^2(BD)-1)(\mathrm{eq}^2(DC)-1)} \\ &\quad = \frac{\cos(\angle C)\cos(\angle B)(\cos^2(\angle B)+\cos^2(\angle C)-1)}{\sin(\angle C)\sin(\angle B)(\cos^2(\angle B)+\cos^2(\angle C))}. \end{aligned} \tag{40}$$

Replacing the right-hand and the left-hand sides of equations (35) and (36) by expressions (37), (39), and (40), for both hyperbolic and spherical cases we obtain the same formula:

$$\mathrm{eq}(a) = \cot(\angle B)\cot(\angle C). \tag{41}$$

It remains to apply formula (31) from the statement of Lemma 3.4 to equation (41) in order to get expression (32). ☐

Corollary 3.6. *Taking into account formulae* (2) *and* (3), *Lemma* 3.5 *can be read as the Pythagoras theorem in both hyperbolic and spherical spaces.*

Lemma 3.7. *Given a triangle* $\triangle ABC$ *as above, then*

$$\frac{\mathrm{eq}(a)}{\mathrm{eq}(b)} = \frac{\mathrm{circ}(a)}{\mathrm{circ}(b)}\cos(\angle C) \quad \textit{and} \quad \frac{\mathrm{eq}(a)}{\mathrm{eq}(c)} = \frac{\mathrm{circ}(a)}{\mathrm{circ}(c)}\cos(\angle B). \tag{42}$$

Proof. By Lemma 3.5 we know that

$$\mathrm{eq}(a) = \cot(\angle B)\cot(\angle C).$$

Also, by Lemma 3.4 it is true that

$$\mathrm{eq}(b) = \frac{\cos(\angle B)}{\sin(\angle C)}.$$

Therefore,

$$\frac{\mathrm{eq}(a)}{\mathrm{eq}(b)} = \frac{\cos(\angle C)}{\sin(\angle B)}. \tag{43}$$

We also know from Lemma 3.3 that

$$\frac{1}{\sin(\angle B)} = \frac{\mathrm{circ}(a)}{\mathrm{circ}(b)}. \tag{44}$$

Substituting (44) into (43), we get the first identity (42). Exchanging roles of b and c and also of B and C, we obtain the second identity (42). ☐

4 Corollaries

Let us consider an arbitrary triangle $\triangle FGH$ in a plane of constant curvature $K \in \mathbb{R}$ with angles $\angle F(<\pi)$, $\angle G(<\pi)$, $\angle H(<\pi)$ at the vertices F, G, and H, and with lengths f, g, and h of the edges GH, HF, and FG, respectively.

Lemma 4.1. *Given a triangle $\triangle FGH$ as above, then*

$$\frac{\mathrm{circ}(f)}{\sin(\angle F)} = \frac{\mathrm{circ}(g)}{\sin(\angle G)} = \frac{\mathrm{circ}(h)}{\sin(\angle H)}. \tag{45}$$

Proof. Let us take a point T_F on the straight line GH such that the segment FT_F is orthogonal to GH, and let t_F denote the length of FT_F. Applying Lemma 3.3 to $\triangle T_F FH$, we get

$$\mathrm{circ}(t_F) = \mathrm{circ}(g)\sin(\angle H)$$

if the point H does not lie between the points G and T_F on the straight line GH, and

$$\mathrm{circ}(t_F) = \mathrm{circ}(g)\sin(\pi - \angle H)$$

if the point H lies between G and T_F on GH. Since $\sin\varphi = \sin(\pi - \varphi)$ for any $\varphi \in \mathbb{R}$, formula

$$\mathrm{circ}(t_F) = \mathrm{circ}(g)\sin(\angle H) \tag{46}$$

holds in both cases.

Similarly, applying Lemma 3.3 to $\triangle T_F FG$, we obtain

$$\operatorname{circ}(t_F) = \operatorname{circ}(h)\sin(\angle G). \tag{47}$$

Mixing (46) and (47), we get

$$\frac{\operatorname{circ}(g)}{\sin(\angle G)} = \frac{\operatorname{circ}(h)}{\sin(\angle H)}. \tag{48}$$

By analogy, if we take a point T_G on the straight line FH such that the segment GT_G is orthogonal to FH and then apply Lemma 3.3 to the triangles $\triangle T_G GH$ $\triangle T_G GF$, we obtain

$$\frac{\operatorname{circ}(f)}{\sin(\angle F)} = \frac{\operatorname{circ}(h)}{\sin(\angle H)}. \tag{49}$$

The equalities (48) and (49) imply (45). □

Corollary 4.2. *Taking into account formulae (2) and (3), Lemma 4.1 can be read as the Law of Sines in both hyperbolic and spherical spaces.*

Lemma 4.3. *Given a triangle $\triangle FGH$ as above, then*

(s) *in the spherical case the following relation holds:*

$$\operatorname{eq}(f) = \operatorname{eq}(g)\operatorname{eq}(h) + \sqrt{(\operatorname{eq}^2(g)-1)(\operatorname{eq}^2(h)-1)}\cos(\angle F); \tag{50}$$

(h) *in the hyperbolic case the following relation holds:*

$$\operatorname{eq}(f) = \operatorname{eq}(g)\operatorname{eq}(h) - \sqrt{(\operatorname{eq}^2(g)-1)(\operatorname{eq}^2(h)-1)}(\cos\angle F). \tag{51}$$

Proof. Remark that Lemma 4.3 reduces to Lemma 3.5 if $\angle F = \frac{\pi}{2}$.

Let us now consider the general situation, that is, when $0 < \angle F < \pi$ and $\angle F \neq \frac{\pi}{2}$.

Take a point K_G on the straight line FH such that the segment GK_G is orthogonal to FH. Let k_G denote the length of GK_G, g_1 denote the length of FK_G, and g_2 denote the length of $K_G H$.

Let us first obtain formula

$$\operatorname{eq}(f) = \operatorname{eq}(g)\operatorname{eq}(h) - \frac{\operatorname{circ}(h)}{\operatorname{circ}(g)}(\operatorname{eq}^2(g)-1)\cos(\angle F). \tag{52}$$

Applying Lemma 3.5 to $\triangle FGK_G$ and to $\triangle HGK_G$, we get

$$\operatorname{eq}(h) = \operatorname{eq}(g_1)\operatorname{eq}(k_G), \tag{53}$$

$$\operatorname{eq}(f) = \operatorname{eq}(g_2)\operatorname{eq}(k_G). \tag{54}$$

Let us study three possible cases.

CASE 1. Assume that $\angle F < \frac{\pi}{2}$, $\angle H \leq \frac{\pi}{2}$. Therefore K_G belongs to the segment FH and $g_1 + g_2 = g$. Substituting $g_2 = g - g_1$ into (54), applying relation (14) to $\mathrm{eq}(g - g_1)$, and taking into account formula (53), we obtain

$$\mathrm{eq}(f) = \mathrm{eq}(g)\,\mathrm{eq}(h) - \frac{\mathrm{circ}(g_1)}{\mathrm{circ}(g)}\frac{\mathrm{eq}(h)}{\mathrm{eq}(g_1)}(\mathrm{eq}^2(g) - 1). \tag{55}$$

Applying Lemma 3.7 to $\triangle FGK_G$ we get

$$\frac{\mathrm{eq}(h)}{\mathrm{eq}(g_1)} = \frac{\mathrm{circ}(h)}{\mathrm{circ}(g_1)}\cos(\angle F). \tag{56}$$

Substituting (56) in (55), we obtain (52).

CASE 2. Assume that $\angle F < \frac{\pi}{2}$, $\angle H > \frac{\pi}{2}$. Therefore H belongs to the segment FK_G and $g_1 = g + g_2$. Substituting $g_2 = g_1 - g$ into (54), recalling that, by (5), $\mathrm{eq}(g_1 - g) = \mathrm{eq}(g - g_1)$, and then repeating step by step what we have already done in Case 1, we obtain (52).

CASE 3. Assume that $\angle F > \frac{\pi}{2}$. Therefore F belongs to the segment $K_G H$ and $g_2 = g + g_1$. Substituting it into (54), applying relation (15) to $\mathrm{eq}(g + g_1)$, and taking into account formula (53) we obtain

$$\mathrm{eq}(f) = \mathrm{eq}(g)\,\mathrm{eq}(h) + \frac{\mathrm{circ}(g_1)}{\mathrm{circ}(g)}\frac{\mathrm{eq}(h)}{\mathrm{eq}(g_1)}(\mathrm{eq}^2(g) - 1). \tag{57}$$

Applying Lemma 3.7 to $\triangle FGK_G$ and considering that $\angle GFK_G = \pi - \angle F$ we get

$$\frac{\mathrm{eq}(h)}{\mathrm{eq}(g_1)} = \frac{\mathrm{circ}(h)}{\mathrm{circ}(g_1)}\cos(\angle GFK_G) = \frac{\mathrm{circ}(h)}{\mathrm{circ}(g_1)}\cos(\pi - \angle F) = -\frac{\mathrm{circ}(h)}{\mathrm{circ}(g_1)}\cos(\angle F). \tag{58}$$

Substituting (58) in (57), we obtain (52).

Thus, formula (52) is proved.

Similarly, if we take a point K_H on the straight line FG such that the segment HK_H is orthogonal to FG and repeat the previous reasoning, we obtain

$$\mathrm{eq}(f) = \mathrm{eq}(g)\,\mathrm{eq}(h) - \frac{\mathrm{circ}(g)}{\mathrm{circ}(h)}(\mathrm{eq}^2(h) - 1)\cos(\angle F). \tag{59}$$

Expressing $\mathrm{eq}(g)\,\mathrm{eq}(h) - \mathrm{eq}(f)$ from (52) and (59), we get

$$\frac{\mathrm{circ}(h)}{\mathrm{circ}(g)}(\mathrm{eq}^2(g) - 1)\cos(\angle F) = \frac{\mathrm{circ}(g)}{\mathrm{circ}(h)}(\mathrm{eq}^2(h) - 1)\cos(\angle F),$$

$$\left(\frac{\mathrm{circ}(h)}{\mathrm{circ}(g)}\right)^2 = \frac{\mathrm{eq}^2(h) - 1}{\mathrm{eq}^2(g) - 1}.$$

Recall that, by definition, $\mathrm{circ}(r) > 0$ for $r > 0$. Therefore,

$$\frac{\mathrm{circ}(h)}{\mathrm{circ}(g)} = \sqrt{\frac{\mathrm{eq}^2(h) - 1}{\mathrm{eq}^2(g) - 1}}. \tag{60}$$

Substituting (60) into (52), we obtain

$$\mathrm{eq}(f) = \mathrm{eq}(g)\,\mathrm{eq}(h) - \sqrt{\frac{\mathrm{eq}^2(h) - 1}{\mathrm{eq}^2(g) - 1}}(\mathrm{eq}^2(g) - 1)\cos(\angle F). \tag{61}$$

As $\mathrm{eq}^2(a) \leq 1$ in the spherical space and $\mathrm{eq}^2(a) \geq 1$ in the hyperbolic space (see formulae (2) and (3)), the identity (61) implies (50) in the spherical case and (51) in the hyperbolic case. ☐

Corollary 4.4. *Taking into account formulae (2) and (3), Lemma 4.3 can be read as the Law of Cosines in both hyperbolic and spherical spaces.*

Acknowledgements. The author of the present paper would like to thank Athanase Papadopoulos for the proposition to study de Tilly's mechanical methods applied to geometric problems and for interesting discussions, and Vincent Alberge for the attentive reading of this text and for his valuable comments and suggestions.

References

[1] J.-M. de Tilly, "Etudes de mécanique abstraite," *Mémoires Acad. Roy. Belg.* XXI (1870), 1–99.
☞ 94, 95

[2] A. Papadopoulos and G. Théret, *La théorie des lignes parallèles de Johann Heinrich Lambert,* Collection Sciences dans l'Histoire, Librairie Scientifique et Technique Albert Blanchard, Paris, 2014.
MR 3241580 Zbl 1304.01004 ☞ 93

[3] A. Papadopoulos and G. Théret, "Hyperbolic geometry in the work of J. H. Lambert," *Gaṇita Bhārātī* 36 (2014), no. 2, 129–155.
MR 3410718 Zbl 1341.01017 ☞ 93

[4] G. Saccheri, *Euclid vindicated from every blemish,* dual Latin–English text, edited and annotated by V. De Risi, Translated from the Latin by G. B. Halsted and L. Allegri, Classic Texts in the Sciences. Birkhäuser/Springer, Cham, 2014.
MR 3243765 Zbl 1303.01003 ☞ 93

[5] J.-D. Voelke, *Renaissance de la géométrie non euclidienne entre 1860 et 1900.* Peter Lang, Bern, 2005.
MR 2143213 Zbl 1078.01014 ☞ 95

Chapter 8

The Gauss–Bonnet theorem and the geometry of surfaces

Son Lam Ho

Contents

1. Introduction 113
2. Geodesic curvature and geodesics 116
3. The Gauss–Bonnet theorem 120
References 123

1 Introduction

The Gauss–Bonnet theorem is one of the most beautiful equations in mathematics. It says that for an orientable closed surface S in Euclidean space $\mathbb{R}^3$, the "total curvature" is equal to 2π the Euler characteristic of S. That is,

$$\iint_S K \, d\sigma = 2\pi\chi(S).$$

The left-hand side is the integral of K over the whole surface S with respect to the area form $d\sigma$; here K is the Gaussian curvature at each point on the surface. The Gaussian curvature is a geometric quantity which changes as you bend, stretch or scale S in $\mathbb{R}^3$. On the other hand, the Euler characteristic $\chi(S)$ is a topological invariant, a quantity that remains constant no matter how you deform the surface. What makes the Gauss–Bonnet formula remarkable is that it elegantly connects these two very different worlds of geometry and topology.

The above *global Gauss–Bonnet theorem* follows directly from the *local Gauss–Bonnet theorem*; we state a simplified version of it below:

Theorem 1.1 (Gauss). *Let S be a surface in $\mathbb{R}^3$ and let $T = \Delta(A, B, C)$ be a geodesic triangle in S, that is, T is a simply connected subset of S which is bounded by three geodesic segments connecting three points A, B, C. Then*

$$\iint_T K\, d\sigma = (\angle A + \angle B + \angle C) - \pi.$$

Here $\angle A, \angle B, \angle C$ are the interior angles of the triangle. Thus, the right-hand side is the angle excess of T. Indeed, since T is not a planar triangle, we should expect its interior to have non-zero total curvature. In that case the theorem proves that the sum of its angles is not π, and how different it is from π depends on curvature and area of T. Important examples are when K is constant, in which case the angle excess of a triangle T is a constant multiple of its area.

Our goal is to give the reader an informal understanding of the concepts related to the above formula without developing the machinery for a rigorous proof which can be found in many textbooks. The references [1] and [4] are good sources for a complete treatment of the subject. In Section 2, we define geodesic curvature of curves on surfaces and demonstrate its relation with angles on a piecewise geodesic curve. In Section 3, we informally explain the Gaussian curvature and provide a heuristic proof of Gauss–Bonnet formulas.

It is enlightening to know the history of how these ideas developed. Given a surface S smoothly embedded in $\mathbb{R}^3$, at each point $p \in S$ there is a tangent plane $T_p S$ containing all tangent vectors, and there is a unit normal vector $\vec{N}_p$ perpendicular to the surface. Since S is orientable there is a coherent and continuous way to choose $\vec{N}_p$ at every point p. Pick a unit tangent vector $\vec{v} \in T_p S$. There is a unique plane $P_{\vec{v}}$ containing both $\vec{N}_p$ and $\vec{v}$, and this plane intersects S along a curve γ containing p. Locally around p we can parametrize this curve $\gamma\colon [-\epsilon, \epsilon] \to S$ such that $\gamma(0) = p$, $\gamma'(0) = \vec{v}$, and such that it is unit speed. Of course the first derivative $\gamma'(t)$ gives us tangent vectors. We can differentiate again to get the second derivative $\gamma''(t)$ of acceleration vectors. Then let us define $\kappa(\vec{v})$ to be the dot product $\vec{N}_p.\gamma''(0)$. This quantity $\kappa(\vec{v})$ will be called the curvature in the $\vec{v}$ direction at p. Since the curve γ is on the plane $P_{\vec{v}}$, it is easy to see that $\kappa(\vec{v}) = \pm\|\gamma''(0)\|$ and the sign depends on whether $\gamma''(0)$ and $\vec{N}_p$ point to the same or opposite direction. The following theorem is proven by Euler [2]; it precedes Gauss' work on surfaces.

Theorem 1.2 (Euler). *Given the above notations, if $\kappa(\vec{v})$ are not all the same as $\vec{v}$ varies in the tangent plane, then there is a unit tangent vector $\vec{v}_1$ such that $k_1 = \kappa(\vec{v}_1)$ is the minimum, and a unit tangent $\vec{v}_2$ such that $k_2 = \kappa(\vec{v}_2)$ is the maximum. Moreover, $\vec{v}_1$ and $\vec{v}_2$ are perpendicular.*

Here k_1, k_2 are called the principal curvatures of S at p, and $\vec{v}_1, \vec{v}_2$ are the corresponding principal directions.

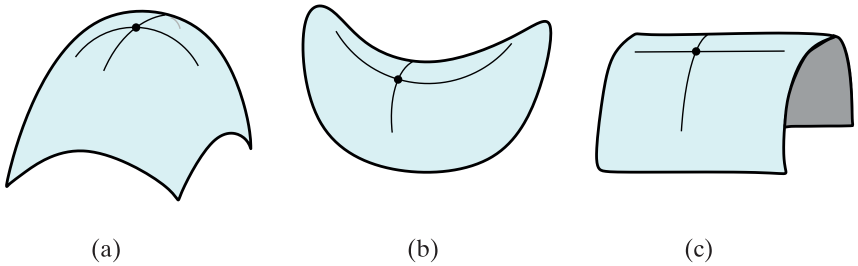
Gauss defined surface curvature as follows [3]. Let $\mathbb{S}^2$ be the unit 2-sphere in $\mathbb{R}^3$. For a smooth surface $S \subset \mathbb{R}^3$, let $\mathcal{G}\colon S \to \mathbb{S}^2$ be the map taking a point $p \in S$ to its unit normal vector $\vec{N}_p$. This is called the Gauss map. We have the differential map $d\mathcal{G}_p\colon T_pS \to T_{\mathcal{G}(p)}\mathbb{S}^2$ where T_pS is the tangent plane to S at p and $T_{\mathcal{G}(p)}\mathbb{S}^2$ is the tangent plane to the unit sphere at $\mathcal{G}(p)$.

Definition 1.3. Let $\vec{v}_1, \vec{v}_2$ be tangent vectors at p. The *Gaussian curvature* at p is defined to be

$$K = \frac{\text{Area of the parallelogram spanned by } \mathcal{G}(\vec{v}_1), \mathcal{G}(\vec{v}_2)}{\text{Area of the parallelogram spanned by } \vec{v}_1, \vec{v}_2}.$$

Of course this is an informal version of a more rigorous definition. We can think of K as the scaling factor of the Gauss map on infinitesimal area near p.

It is then showed that the curvature K at p equals the product of the principal curvatures, that is, $K = k_1k_2$. This result gives us a geometric understanding of curvature of surfaces. At a point on S, K is positive if its principal curvatures are of the same sign, which is when the principal tangent curves bend in the same direction at p; K is negative when the principal tangent curves bend in opposite directions; and K is zero when one of the principal tangent curves is flat at p. Figure 1 demonstrates the local picture around points of positive, negative, and zero curvatures.



(a) (b) (c)

Figure 1

When a surface is smoothly embedded in $\mathbb{R}^3$, it inherits an infinitesimal metric which is given in coordinates by the 1st fundamental form. This metric, later known as the Riemannian metric—a notion of distance on the surface—exists without the need of embedding the surface into 3-space. Gauss formulated a more sophisticated description of K in terms of the 1st and 2nd fundamental forms, which in turn can be computed purely using the 1st fundamental form. This result is called Gauss' *Theorema egregium* [3], translated by "remarkable theorem." Indeed the theorem shows that the curvature K depends only on the Riemannian metric of the surface, hence it is an *intrinsic* property of the surface and it can be computed purely by measurements of distances and angles on the surface. This result and the works of Bernhard Riemann that follows mark the beginning of Differential Geometry, a very important and beautiful area of mathematics.

Assuming the *Theorema Egregium*, in Section 3, we will use an informal intrinsic definition of the Gaussian curvature to prove the local Gauss–Bonnet theorem.

2 Geodesic curvature and geodesics

We will now define the notion of geodesics which will be important to understand curves on surfaces. Let S be a smooth surface in 3-space. Consider a smooth curve $\gamma\colon [a,b] \to S$ of unit speed, that is, $\|\gamma'(t)\| = 1$. At the point $\gamma(t)$, the curve's tangent vector is given by $\gamma'(t)$ and its acceleration vector is $\gamma''(t)$. Since the curve is unit speed, the acceleration vector is perpendicular to the tangent vector. (One can see this by moving the base point of vector $\gamma'(t)$ to 0, so that $\gamma'(t)$ becomes a path on the unit sphere, which makes the vector $\gamma''(t)$ tangent to this path at the point $\gamma'(t)$, hence $\gamma'(t) \perp \gamma''(t)$.) Now let $(\gamma'(t))^{\perp}$ be the plane of all vectors perpendicular to $\gamma'(t)$.

We have a natural orthonormal basis for $(\gamma'(t))^{\perp}$ given by $\{\vec{N}_{\gamma(t)}, \vec{N}_{\gamma(t)} \times \gamma'(t)\}$. Recall that $\vec{N}_{\gamma(t)}$ is the unit normal vector to the surface and $\vec{N}_{\gamma(t)} \times \gamma'(t)$ is the cross product which is perpendicular to both $\vec{N}_{\gamma(t)}$ and $\gamma'(t)$. Since $\gamma''(t) \in (\gamma'(t))^{\perp} = \operatorname{span}\{\vec{N}_{\gamma(t)}, \vec{N}_{\gamma(t)} \times \gamma'(t)\}$, we have a unique decomposition

$$\gamma''(t) = k_n(t)\vec{N}_{\gamma(t)} + k_g(t)(\vec{N}_{\gamma(t)} \times \gamma'(t))$$

for $k_n(t), k_g(t) \in \mathbb{R}$.

Definition 2.1. We define the *geodesic curvature* of the curve γ at point $p = \gamma(t)$ to be $k_g(t)$.

Definition 2.2. A geodesic on S is a unit speed curve whose geodesic curvature is 0 everywhere along the curve.

The geodesic curvature is a real valued quantity measuring how much the curve is veering left or right at a given point, left and right is distinguished by the sign of $k_g(t)$. The curve is geodesic when informally speaking it is always going straight, and more precisely, it is geodesic when the acceleration vector $\gamma''(t)$ is either 0 or perpendicular to the surface.

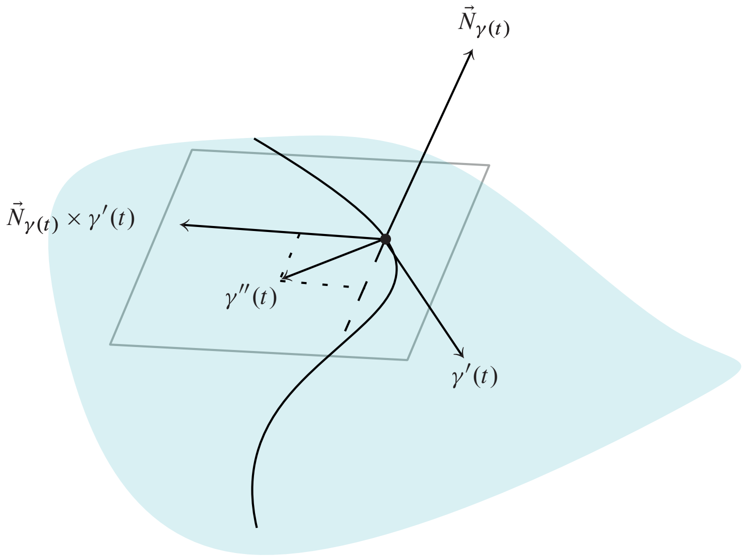


Figure 2

Using this definition one can easily see that geodesics on a plane are straight lines, and circles on a plane has constant geodesic curvature inversely proportional to the circle's radius. Indeed, suppose γ is a unit speed parametrized planar circle of radius r, so $\gamma\colon [0, 2\pi r] \to \mathbb{R}^2$ given by $\gamma(t) = \big(r \cos\left(\frac{t}{r}\right), r \sin\left(\frac{t}{r}\right)\big)$. We can check that

$$\gamma'(t) = \left(-\sin\left(\frac{t}{r}\right), \cos\left(\frac{t}{r}\right)\right) \quad \text{and} \quad \gamma''(t) = \frac{1}{r}\left(-\cos\left(\frac{t}{r}\right), -\sin\left(\frac{t}{r}\right)\right).$$

We can choose $\vec{N}_{\gamma(t)}$ so that $\vec{N}_{\gamma(t)} \times \gamma'(t) = \big(-\cos\left(\frac{t}{r}\right), -\sin\left(\frac{t}{r}\right)\big)$ thus the geodesic curvature is $k_g(t) = \frac{1}{r}$.

Note that for the case of planar circles, the acceleration vector $\gamma''(t)$ always points towards the center of the circle. Using this fact one can see that great circles on a sphere, i.e. circles which are intersection of the sphere with a plane through the center, are geodesics of the sphere.

We defined the geodesic curvature making full use of the ambient space $\mathbb{R}^3$. However, it can be defined intrinsically in a way that depends on the introduction of the unique *Levi-Civita connection* determined by the Riemannian metric. We will not develop this general machinery here.

For the remaining part of this section we will explore how geodesic curvature is related to angles. Given a smooth unit speed arc $\gamma\colon [a,b] \to S$, we can integrate the geodesic curvature with respect to arc-length:

$$\int_{\gamma} k_g\, ds = \int_a^b k_g(t)dt.$$

Think of this quantity as the *total geodesic curvature* of the arc γ. Now suppose that our surface is the $\mathbb{R}^2$ plane, we will compute the total geodesic curvature of γ as the change in direction of the tangent vector $\gamma'(t)$ as t goes from a to b. Since γ is of unit speed, we can think of $\gamma'(t)$ as a parametrized path on the unit circle $\mathbb{S}^1$. For simplicity let us suppose that $k_g(t) \geq 0$ on the whole curve γ. (For curves with k_g changing sign, we can always decompose into a concatenation of sub-arcs so that on each arc k_g is either non-positive or non-negative.) Then we have

$$\int_a^b k_g(t)dt = \int_a^b \|\gamma''(t)\|dt.$$

The right-hand side is exactly the definition of arclength for the path $\gamma'(t)$ which starts at $\gamma'(a)$ and ends at $\gamma'(b)$. Since the arc $\gamma'(t)$ is on the unit circle, its length is $\Theta + k2\pi$ where Θ is the angle between the $\gamma'(a)$ and $\gamma'(b)$ as vectors based at 0, and $k \in \mathbb{Z}$ records how many times the path $\gamma'(t)$ winds around the unit circle.

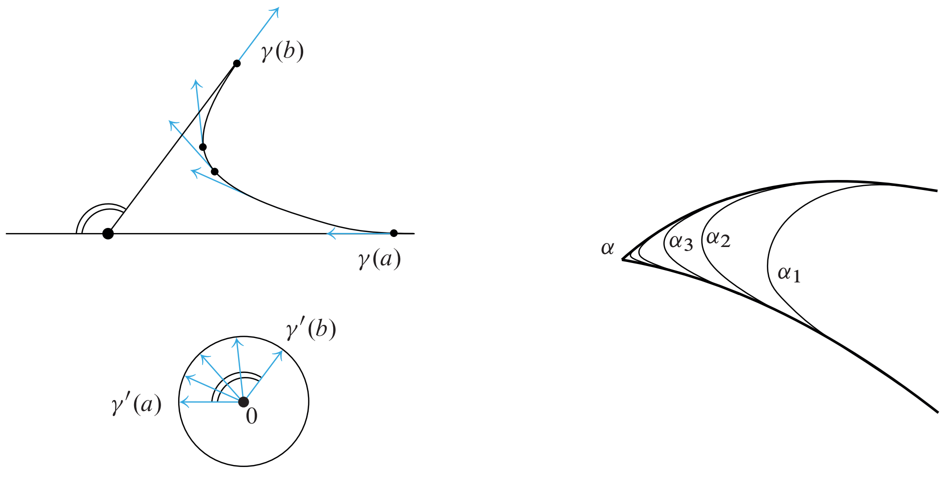
Remark 2.3. The total geodesic curvature of a planar curve γ is equal to the total amount of turning (measured by angle) of the tangent vector $\gamma'(t)$.

We also have the following claim which will be justified below.

Remark 2.4. We can think of an angle as infinitely concentrated geodesic curvature.

Consider two infinite unit speed geodesic rays $a_1, a_2\colon [0,\infty) \to S$ starting at the same point $p_0 = a_1(0) = a_2(0)$. Let $\Theta \in [0,\pi]$ be the angle between $-a_1'(0)$ and $a_2'(0)$. We let $\alpha\colon \mathbb{R} \to S$ be the piecewise geodesic curve so that $\alpha(s) = a_2(s)$ for $s \in [0,\infty)$, and $\alpha(s) = a_1(-s)$ for $s \in (-\infty, 0]$. So Θ is the *exterior angle* at the vertex $p_0 = \alpha(0)$.

One can construct a sequence of smooth curves $\alpha_1, \alpha_2, \ldots$ approximating α by smoothing out the angle as follows. Let $U_1 \supset U_2 \supset \cdots$ be a sequence of open discs all centered at p_0 such that the radius of U_i approaches 0 as $i \to \infty$. We let $\bar{U_i} = S \backslash U_i$ be their complements. We can construct smooth unit speed curves α_i so



(a) (b)

Figure 3. (a) Illustration for Remark 2.3; (b) illustration for Remark 2.4

that up to a reparametrization it agrees with α in $\bar{U}_i$. In other words, α_i agrees with α everywhere except in a small disc U_i around the vertex p_0. Integrating the geodesic curvature of these approximating curves we have

$$\int_{\alpha_i} k_g \, ds = \int_{\alpha_i \cap U_i} k_g \, ds + \int_{\alpha_i \cap \bar{U}_i} k_g \, ds = \int_{\alpha_i \cap U_i} k_g \, ds$$

since $\alpha_i \cap \bar{U}_i$ is where α_i agrees with α which have 0 geodesic curvature. Now since the U_i's are getting smaller, the geometry on U_i gets closer to planar geometry where Remark 2.3 applies. Along $\alpha_i \cap U_i$, the total turning angle for the tangent vectors $\alpha_i'(t)$ approaches Θ, so $\int_{\alpha_i \cap U_i} k_g \, ds \to \Theta$, and therefore

$$\int_{\alpha} k_g \, ds = \lim_{i \to \infty} \int_{\alpha_i} k_g \, ds = \Theta.$$

It makes sense for α, a curve which is geodesic except at one point, to still have a total geodesic curvature measure-concentrated at p_0. Indeed there are points on $\alpha_i \cap U_i$ where the geodesic curvature approaches infinity as i increases.

The full statement of the local Gauss–Bonnet theorem is as follows.

Theorem 2.5 (Gauss). *Let S be a surface in $\mathbb{R}^3$ and let D be an open simply connected region on S whose boundary ∂D is smooth except at n vertices where the exterior angles at these vertices are $\theta_1, \ldots, \theta_n$.*

$$\iint\limits_D K\, d\sigma = 2\pi - \left(\sum_{i=1}^{n} \theta_i + \int\limits_{\partial D} k_g\, ds \right)$$

where $\int_{\partial D} k_g\, ds$ is computed as a sum of integrals over the smooth segments of ∂D.

If $n = 0$, that is, if ∂D is smooth with no vertices, then the equation reduces to $\iint_D K\, d\sigma = 2\pi - \int_{\partial D} k_g\, ds$. This simplified case can be proven by working in an orthogonal uv-coordinate patch and writing both K and k_g in terms of the first fundamental form quantities E, G and partial derivatives. The formula follows after applying the Gauss–Green theorem. This is the rigorous approach as seen in [1].

Now suppose that ∂D has $n \geq 1$ vertices. We can approximate D by a sequence of open domains D_i with smooth boundary ∂D_i agreeing with ∂D on almost all the smooth parts and replacing the angles with short arcs of large geodesic curvature, just like our discussion following Remark 2.4. So as ∂D_i approach ∂D we have the *total measure* of geodesic curvature:

$$\lim_{D_i \to D} \int\limits_{\partial D_i} k_g\, ds = \sum_{i=1}^{n} \theta_i + \int\limits_{\partial D} k_g\, ds.$$

Moreover,

$$\lim_{D_i \to D} \iint\limits_{D_i} K\, d\sigma = \iint\limits_{D} K\, d\sigma.$$

Thus theorem 2.5 follows from the case where ∂D is smooth.

3 The Gauss–Bonnet theorem

One thing we did not explain was how in a coordinate patch, the Gaussian curvature can be expressed as the exterior derivative of geodesic curvature—an essential step to prove the Gauss–Bonnet theorem. To explain this we will define the curvature in a different way, hoping to informally justify the Gauss–Bonnet theorem with a different perspective.

Suppose P is a geodesic polygon on S with k vertices. We only consider the case when P bounds a simply connected interior. Let $\phi_1, \dots, \phi_k$ be the interior angles of P. We can consider following quantity representing the *angle excess of* P:

$$\delta(P) := \sum_{i=1}^{k} \phi_i - (k-2)\pi = 2\pi - \sum_{i=1}^{k} (\pi - \phi_i).$$

Note that $(k-2)\pi$ is the sum of interior angles of a Euclidean k-gon, and 2π is the sum of exterior angles. If we "add" a triangle T next to P (see Figure 4), we obtain a new geodesic polygon Q of $k+1$ vertices. Let $\alpha_1, \alpha_2, \alpha_3$ be the interior angles of T and $\beta_1, \dots, \beta_{k+1}$ be the interior angles of Q.

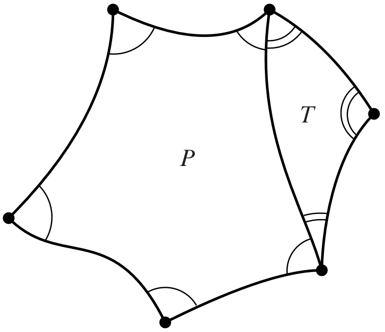


Figure 4

Remark 3.1. Angle excess is *additive*, that is, the angle excess of the new polygon Q is the sum of angle excesses of P and T:

$$\delta(Q) = \sum_{i=1}^{k+1} \beta_i - (k-1)\pi = \sum_{i=1}^{k} \phi_i - (k-2)\pi + \sum_{i=1}^{3} \alpha_i - \pi = \delta(P) + \delta(T).$$

Let p be a point on the smooth surface $S \subset \mathbb{R}^3$. Let $T = \Delta(A, B, C)$ be a geodesic triangle on S. We now define the curvature at p to be

$$K(p) = \lim_{T \to p} \frac{\delta(T)}{\mathrm{Area}(T)} = \lim_{T \to p} \frac{(\angle A + \angle B + \angle C) - \pi}{\mathrm{Area}(T)}.$$

The limit is taken over smaller and smaller geodesic triangles T containing p in a way such that the vertices A, B, C approach p. We will take for granted that this limit exists not depending on how the vertices approach p, and that it varies continuously as p varies in S. This is an intrinsic definition of the curvature because measurements of angles and area are taken on the surface S without refering to how S is embedded in $\mathbb{R}^3$.

Theorem 3.2 (Gauss). *Let S be a surface in $\mathbb{R}^3$ and let $T = \Delta(A, B, C)$ be a geodesic triangle in S, that is, T is a simply connected subset of S which is bounded by three geodesic segments connecting three points A, B, C. Then*

$$\iint\limits_T K\, d\sigma = (\angle A + \angle B + \angle C) - \pi.$$

Informal proof. Since the closure of T is compact, for any $\epsilon > 0$, we can divide T into n sub-triangles $T_{\epsilon,1}, \dots, T_{\epsilon,n}$ such that each $T_{\epsilon,i}$ is of diameter less than ϵ. Here n depends on ϵ by this relation $n \approx \left(\frac{\text{diameter}(T)}{\epsilon}\right)^2$.

We choose a point $p_{\epsilon,i}$ in each $T_{\epsilon,i}$. Applying Riemann integration using the sub-triangles $T_{\epsilon,i}$ we have the following heuristic argument

$$\begin{aligned}\iint\limits_T K\, d\sigma &= \lim_{\epsilon\to 0} \sum_{i=1}^{n} \big(K(p_{\epsilon,i})\, \text{Area}(T_{\epsilon,i})\big)\\ &= \lim_{\epsilon\to 0} \sum_{i=1}^{n} \frac{\delta(T_{\epsilon,i})}{\text{Area}(T_{\epsilon,i})}\, \text{Area}(T_{\epsilon,i})\\ &= \delta(T) = (\angle A + \angle B + \angle C) - \pi \qquad \text{(by Remark 3.1).}\end{aligned}$$

Therefore we obtain the local Gauss–Bonnet formula. The above argument works not only for triangles but for geodesic polygons in general. ☐

We have the global Gauss–Bonnet immediately after the following elegant proof.

Theorem 3.3 (Gauss). *Let S be a smooth closed surface then*

$$\iint\limits_S K\, d\sigma = 2\pi\chi(S).$$

where $\chi(S)$ is the Euler characteristic.

Proof. We give S a triangulation by geodesic triangles T_i. Suppose that the number of vertices, edges, and faces are V, E, F respectively. We have $\chi(S) = V - E + F$ by Euler's famous formula. This is a topological invariant of the surface S not depending on the triangulation. Since each triangle has three edges, we have $3F$ equals the total number of edges double counting each edge twice, because each edge belongs to two faces. So $3F = 2E$, which implies $\chi(S) = V - E + F = V - \frac{F}{2}$.

Let $\beta_1^{(i)}, \beta_2^{(i)}, \beta_3^{(i)}$ be the interior angles of triangle T_i. Now applying Theorem 3.2 we have

$$\begin{aligned}\iint_S K\,d\sigma &= \sum_{i=1}^{F} \iint_{T_i} K\,d\sigma = \sum_{i=1}^{F} \delta(T_i)\\ &= \sum_{i=1}^{F} (\beta_1^{(i)} + \beta_2^{(i)} + \beta_3^{(i)} - \pi)\\ &= \sum_{i=1}^{F} (\beta_1^{(i)} + \beta_2^{(i)} + \beta_3^{(i)}) - F\pi\\ &= 2\pi V - F\pi\\ &= 2\pi\chi(S). \qquad \square\end{aligned}$$

Theorem 3.2 can be applied similarly to obtain a Gauss–Bonnet formula for surfaces with cone points.

Corollary 3.4. *Let S be a surface with k cone points and cone angles $\theta_1, \dots, \theta_k$, then*

$$\iint_S K\,d\sigma = 2\pi\chi(S) + \sum_{i=1}^{k} (\theta_i - 2\pi).$$

References

[1] M. do Carmo, *Differential geometry of curves and surfaces,* translated from the Portuguese, Prentice-Hall, Englewood Cliffs, N.J., 1976.
MR 0394451 Zbl 0326.53001 ☞ 114, 120

[2] L. Euler, “Recherches sur la courbure des surfaces,” *Memoires de l’academie des sciences de Berlin* 16 (1767), 119–143.
☞ 114

[3] C. F. Gauss, *Disquisitiones generales circa superficies curvas,* Typis Dieterichianis, Göttingen, 1828.
`https://archive.org/details/disquisitionesg00gausgoog`
☞ 115

[4] M. Spivak, *A comprehensive introduction to differential geometry,* Vol. II, 3rd ed., Publish or Perish, Houston, TX, 1999.
MR 0532831 Zbl 1213.53001 ☞ 114

Chapter 9

On the non-existence of a perfect map from the 2-sphere to the Euclidean plane

Charalampos Charitos and Ioannis Papadoperakis

Let E^2 be the Euclidean plane and let $S^2 \subset E^3$ be the 2-dimensional unit sphere equipped with its usual metric. With respect to this metric, which will be referred to as *the angular metric*, the distance of two points $p, q \in S^2$ is the length of the smallest arc of the unit circle $C \subset S^2$, with $p, q \in C$.

A smooth map f from S^2 or from a part of S^2 to E^2 will be called *perfect* if for each $p \in S^2$ there is a neighborhood $U(p)$ of p in S^2 such that the restriction of f on $U(p)$ preserves distances infinitesimally along the meridians and the parallels of S^2 and, furthermore, f preserves angles between meridians and parallels, where these angles are measured infinitesimally as it will be explained before Lemma 2.

A local isometry $f\colon X \to Y$ between two metric spaces X and Y is a map which preserves distances locally; that is, for each $p \in X$ there exists a neighborhood $U(p)$ of p in X such that the restriction of f on $U(p)$ preserves distances.

In this chapter, we rewrite a proof of a result of Euler which affirms that a perfect map does not exist. Notice that the non-existence of perfect maps was known from antiquity although a precise definition was never given. A perfect map was understood as a map from the sphere or from a part of the sphere into the Euclidean plane which respects the basic features of the terrestrial globe, namely, the size, the shape and the area. Consequently, it was a common assumption that it was impossible to find a single map preserving at the same time all these features; in other words, the distortion of geographical maps was an unavoidable problem and it was assumed that cartographers should decide which properties to preserve at the expense of others (see [4], Chapter 2). Even in Euler's paper [1] the definition of a perfect map is not explicitly written and one has to read between the lines. In the present paper, the proof of non-existence of a perfect map is along the lines of Euler's argument. It is written in modern language and it is independent of certain geometrical considerations present in Euler's original paper. For instance, Euler uses in an essential way the fact that the vertices p, q, s, r of a quadrilateral of S^2, whose sides consist of arcs belonging to meridians or to parallels, are mapped via a perfect map, to points P, Q, S, R respectively, which form a quadrilateral of E^2.

Remark 1. It is worth noting that Euler's theorem was published in 1777, the year that Gauss was born, and is part of a longer paper [1] concerning the problem of cartography and the construction of plane geographical maps. Gauss proved in 1828 his famous "Theorema Egregium" which states that, if S_1 and S_2 are two surfaces and p, q are points in S_1 and S_2 respectively such that the curvature at p is different from the curvature at q, then there is no local isometry f from S_1 to S_2 with $f(p) = q$.

Using the Riemannian metric of the unit sphere S^2 (see formula (1) below) we can prove that a perfect map from S^2 to E^2 is a local isometry. Therefore, Euler's theorem may be considered as a special case of the theorema egregium. On the other hand, one may show directly that a local isometry from S^2 to E^2 does not exist because there is no way to map a spherical triangle isometrically onto a Euclidean triangle. Euler knew this result. For example in [1] he writes that the non-existence of a perfect map $f\colon S^2 \to E^2$ is a well-known result. However, he gave in [1] a new proof for the non-existence of f through admirable computations, avoiding spherical trigonometry. The basic idea of the proof is to translate geometrical conditions to a system of differential equations which does not have a solution. Euler's method is very fruitful and can be applied for the proof of similar results. For example, it is not hard to prove the following generalization of Theorem 1: there is no map f from S^2 to E^2 which preserves distances infinitesimally along the meridians and the parallels of S^2 and which sends the angles between meridians and parallels to a constant angle.

Finally, it is shown by the authors that the hypothesis that a map f preserves distances infinitesimally along the meridians is weaker than that of preserving the distance between any two points on them. This will be done by means of Proposition 5.

Below we will present first a proof of the following theorem, proved in [1] (see [2] for the translation of [1] in English).

Theorem 1. *There does not exist any smooth map Φ from a neighborhood of the sphere S^2 to a neighborhood of the Euclidean plane E^2 which is perfect.*

We consider a parametrization of the Eastern hemisphere S_0 of S^2 given by the map

$$p\colon \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \times \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \longrightarrow S_0$$

with

$$p(t, u) = (\cos(t)\cos(u), \sin(t)\cos(u), \sin(u)).$$

Thus, the u-curves are semicircles lying on a meridian of S^2 and are geodesics of S^2, while the t-curves are semicircles parallel to the equator of S^2 and are not geodesics.

With respect to the coordinates (t, u), the Riemannian metric on S^2 has the form

$$ds = \sqrt{du^2 + dt^2 \cos^2 u}. \tag{1}$$

This formula was obtained by Euler as a corollary of one of his constitutive principles of spherical geometry. This principle is contained in [3] and states that infinitesimally, spherical geometry behaves like Euclidean geometry. Obviously, the distance between two points $p, q \in S^2$ defined by means of (1), coincides with the angular distance. The Riemannian metric ds will not be used in the calculations below.

We denote by $|p_1 - p_2|$ the angular distance between the points $p_1, p_2 \in S^2$ and by $\|P_1 - P_2\|$ the Euclidean distance between the points $P_1, P_2 \in E^2$.

Let $t, dt, t + dt, u, du, u + du \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, $du \neq 0$, $dt \neq 0$, where dt, du are considered as finite variations of t, u respectively. Then, see Figure 1, we set for short

$$p(t, u) = p, \quad p(t, u + du) = q, \quad p(t + dt, u) = r.$$

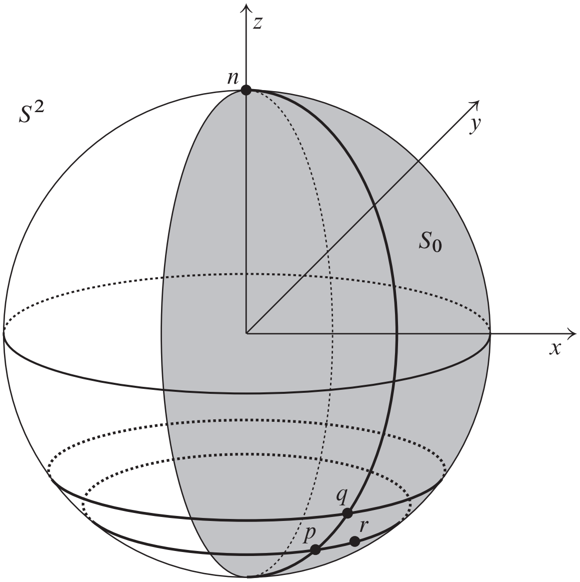


Figure 1. A parametrization of the Eastern hemisphere S_0 with three points, $p(t, u) = p$, $p(t, u + du) = q$, and $p(t + dt, u) = r$

Now, the parallels and the meridians of S^2 intersect each other perpendicularly, as curves. But since the parallels are not geodesics of S^2, the angle $\widehat{qpr}$ of the spherical triangle pqr is not right. For this reason, Euler measures the angles between meridians and parallels infinitesimally. That is, the angle between the meridian and the parallel at the point p is by definition equal to

$$\lim_{dt \to 0} \widehat{qpr}.$$

The following lemma asserts that this angle is right and this explains the condition imposed in the definition of a perfect map, where the right angles between meridians and parallels are preserved.

Lemma 2. *We have*

$$\lim_{dt \to 0} \widehat{qpr} = \frac{\pi}{2}.$$

Proof. Let $n = (0, 0, 1)$ be the north pole of S^2. Considering on S^2 the spherical triangle T, whose vertices are n, p and r (see Figure 1), by Girard's theorem, the area E of T is equal to

$$E = \widehat{nrp} + \widehat{rpn} + \widehat{pnr} - \pi. \tag{2}$$

We also have

$$\widehat{nrp} = \widehat{rpn} = \widehat{rpq}. \tag{3}$$

On the other hand, $\widehat{pnr} = |dt|$ and $\lim_{dt \to 0} E = 0$. Therefore, from relations (2) and (3) we obtain

$$\lim_{dt \to 0} \widehat{qpr} = \lim_{dt \to 0} \frac{1}{2}(E + \pi - \widehat{pnr}) = \frac{\pi}{2}. \qquad \square$$

We also need the following lemma.

Lemma 3. *We have*

$$\sin\left(\frac{1}{2}|r - p|\right) = \left|\sin\left(\frac{1}{2}dt\right)\right| \cos(u).$$

Proof. The points p, r belong to a parallel circle C of radius $\cos u$. Therefore these points define a chord in C of length $2\left|\sin\left(\frac{1}{2}dt\right)\right|\cos u$. On the other hand, the points p, r belong to a geodesic circle C_g of S^2, i.e. a great circle of S^2, and thus the chord joining them has length equal to $2\sin\left(\frac{1}{2}|r - p|\right)$. Therefore our equality follows.

□

Lemma 4. *We have*

$$\lim_{dt\to 0}\frac{|r-p|}{|dt|}=\cos(u).$$

Proof. From Lemma 3 we have

$$\sin\left(\frac{1}{2}|r-p|\right)=\left|\sin\left(\frac{1}{2}dt\right)\right|\cos(u).$$

Therefore,

$$\begin{aligned}\lim_{dt\to 0}\frac{|r-p|}{|dt|}&=\lim_{dt\to 0}\frac{\frac{1}{2}|r-p|}{\sin\left(\frac{1}{2}|r-p|\right)}\frac{\left|\sin\left(\frac{1}{2}dt\right)\right|\cos(u)}{\frac{1}{2}|dt|}\\&=\lim_{w\to 0}\frac{\frac{1}{2}w}{\sin\left(\frac{1}{2}w\right)}\lim_{dt\to 0}\frac{\left|\sin\left(\frac{1}{2}dt\right)\right|}{\frac{1}{2}|dt|}\cos(u)\\&=\cos(u).\end{aligned}$$

□

Remark 2. Lemma 4 and the fact that spherical geometry is infinitesimally Euclidean implies formula (1) for the Riemannian metric ds on S^2.

Proof of Theorem 1. We assume that Φ is a perfect map defined in a neighborhood of p and we will get a contradiction.

We set

$$\Phi(p(t,u))=P(t,u)\quad\text{and}\quad P(t,u)=(x(t,u),y(t,u)).$$

For short, we set

$$P(t,u)=P,\quad P(t,u+du)=Q,\quad P(t+dt,u)=R,$$
$$x(t,u)=x,\quad y(t,u)=y.$$

The existence of the perfect map Φ implies that

$$\lim_{du\to 0}\frac{|q-p|}{|du|}=\lim_{du\to 0}\frac{\|Q-P\|}{|du|},\tag{3}$$

$$\lim_{dt\to 0}\frac{|r-p|}{|dt|}=\lim_{dt\to 0}\frac{\|R-P\|}{|dt|},\tag{4}$$

$$\lim_{dt\to 0}\widehat{qpr}=\lim_{dt\to 0}\widehat{QPR}.\tag{5}$$

On the other hand we have

$$\lim_{du\to 0}\frac{|q-p|}{|du|}=1,$$

$$\lim_{du\to 0}\frac{\|Q-P\|}{|du|}=\sqrt{\left(\frac{\partial x}{\partial u}\right)^2+\left(\frac{\partial y}{\partial u}\right)^2},$$

$$\lim_{dt\to 0}\frac{\|R-P\|}{|dt|}=\sqrt{\left(\frac{\partial x}{\partial t}\right)^2+\left(\frac{\partial y}{\partial t}\right)^2}.$$

From Lemma 4,

$$\lim_{dt\to 0}\frac{|r-p|}{|dt|}=\cos(u),$$

and from Lemma 2 and relation (5)

$$\lim_{dt\to 0}\cos(\widehat{QPR})=0.$$

This implies that

$$\begin{aligned}
0 &= \lim_{dt\to 0}\left|\frac{\langle\overrightarrow{PQ},\overrightarrow{PR}\rangle}{\|P-Q\|\|P-R\|}\right|\\
&= \lim_{du\to 0}\lim_{dt\to 0}\left|\frac{\langle\overrightarrow{PQ},\overrightarrow{PR}\rangle}{\|P-Q\|\|P-R\|}\right|\\
&= \lim_{du\to 0}\lim_{dt\to 0}\frac{|\langle(P(t,u+du)-P),(P(t+dt,u)-P)\rangle|}{\|P(t,u+du)-P\|\cdot\|P(t+dt,t)-P\|}\\
&= \lim_{du\to 0}\lim_{dt\to 0}\frac{\left|\left\langle\frac{(x(t,u+du)-x,\ y(t,u+du)-y)}{du},\frac{(x(t+dt,u)-x,\ y(t+dt,u)-y)}{dt}\right\rangle\right|}{\left\|\frac{(x(t,u+du)-x,\ y(t,u+du)-y)}{du}\right\|\left\|\frac{(x(t+dt,u)-x,\ y(t+dt,u)-y)}{dt}\right\|}\\
&= \frac{\left|\left\langle\left(\frac{\partial x}{\partial u},\frac{\partial y}{\partial u}\right),\left(\frac{\partial x}{\partial t},\frac{\partial y}{\partial t}\right)\right\rangle\right|}{\left\|\left(\frac{\partial x}{\partial u},\frac{\partial y}{\partial u}\right)\right\|\left\|\left(\frac{\partial x}{\partial t},\frac{\partial y}{\partial t}\right)\right\|}\\
&= \frac{\left|\left(\frac{\partial x}{\partial u}\frac{\partial x}{\partial t}+\frac{\partial y}{\partial u}\frac{\partial y}{\partial t}\right)\right|}{\sqrt{\left(\frac{\partial x}{\partial u}\right)^2+\left(\frac{\partial y}{\partial u}\right)^2}\sqrt{\left(\frac{\partial x}{\partial t}\right)^2+\left(\frac{\partial y}{\partial t}\right)^2}},
\end{aligned}$$

where $\langle\cdot,\cdot\rangle$ denotes the Euclidean dot product.

Hence, from (3)–(5), we obtain the following relations:

$$\sqrt{\left(\frac{\partial x}{\partial u}\right)^2+\left(\frac{\partial y}{\partial u}\right)^2}=1, \tag{6}$$

$$\sqrt{\left(\frac{\partial x}{\partial t}\right)^2+\left(\frac{\partial y}{\partial t}\right)^2}=\cos(u), \tag{7}$$

$$\frac{\partial x}{\partial u}\frac{\partial x}{\partial t}+\frac{\partial y}{\partial u}\frac{\partial y}{\partial t}=0. \tag{8}$$

Now, from relations (6) and (7), there exists (ϕ,ω) so that

$$\left(\frac{\partial x}{\partial u},\frac{\partial y}{\partial u}\right)=(\cos(\phi),\sin(\phi)),$$

$$\left(\frac{\partial x}{\partial t},\frac{\partial y}{\partial t}\right)=(\cos(u)\cos(\omega),\cos(u)\sin(\omega)).$$

Using relation (8), we deduce that $\omega=\frac{\pi}{2}+\phi$ or $\phi=\frac{3\pi}{2}+\phi$. If we assume that $\omega=\frac{\pi}{2}+\phi$ (the other case is treated similarly), we obtain

$$\left(\frac{\partial x}{\partial u},\frac{\partial y}{\partial u}\right)=(\cos(\phi),\sin(\phi)), \tag{9}$$

$$\left(\frac{\partial x}{\partial t},\frac{\partial y}{\partial t}\right)=(-\cos(u)\sin(\phi),\cos(u)\cos(\phi)). \tag{10}$$

Therefore, since

$$\frac{\partial^2}{\partial u\partial t}=\frac{\partial^2}{\partial t\partial u}$$

Relations (9) and (10) imply that

$$-\sin(\phi)\frac{\partial\phi}{\partial t}=-\cos(\phi)\frac{\partial\phi}{\partial u}\cos(u)+\sin(\phi)\sin(u), \tag{11}$$

$$\cos(\phi)\frac{\partial\phi}{\partial t}=-\sin(\phi)\frac{\partial\phi}{\partial u}\cos(u)-\cos(\phi)\sin(u). \tag{12}$$

Solving the system of equations (11) and (12) we obtain that

$$\frac{\partial\phi}{\partial u}=0 \quad\text{and}\quad \frac{\partial\phi}{\partial t}=-\sin u.$$

Hence,

$$0=\frac{\partial^2\phi}{\partial t\partial u}=\frac{\partial^2\phi}{\partial u\partial t}=-\cos u,$$

which gives a contradiction. □

Now in the spirit of Euler's method we will prove the following proposition.

Proposition 5. *There does not exist a smooth map Ψ from a neighborhood U of S^2 to E^2 which preserves infinitesimally the distance along the meridians and the parallels of S^2 and which sends the meridional arcs of $U \cap S^2$ to straight lines of E^2.*

Proof. We assume that such a map Ψ exists and we will get a contradiction. Let $\Psi(t,u) = (x(t,u), y(t,u))$, where (t,u) are the parameters introduced above. Since Ψ preserves infinitesimally the distance along the meridians and the parallels of S^2, we have from relations (6) and (7) in the proof of Theorem 1 that there are variables ϕ and ω which are functions of t, u such that

$$\left(\frac{\partial x}{\partial u}, \frac{\partial y}{\partial u}\right) = (\cos(\phi), \sin(\phi)),$$

$$\left(\frac{\partial x}{\partial t}, \frac{\partial y}{\partial t}\right) = (\cos(u)\cos(\omega), \cos(u)\sin(\omega)).$$

Since

$$\frac{\partial^2}{\partial u \partial t} = \frac{\partial^2}{\partial t \partial u},$$

we have

$$-\sin(\phi)\frac{\partial \phi}{\partial t} = -\sin(u)\cos(\omega) - \cos(u)\sin(\omega)\frac{\partial \omega}{\partial u}, \tag{13}$$

$$\cos(\phi)\frac{\partial \phi}{\partial t} = -\sin(u)\sin(\omega) + \cos(u)\cos(\omega)\frac{\partial \omega}{\partial u}. \tag{14}$$

Multiplying (13) with $\cos(\omega)$ and (14) with $\sin(\omega)$ and adding, we have

$$\begin{aligned} (-\sin(\phi)\cos(\omega) + \cos(\phi)\sin(\omega))\frac{\partial \phi}{\partial t} &= -\sin(u)\cos^2(\omega) - \sin(u)\sin^2(\omega) \\ \iff \sin(\phi - \omega)\frac{\partial \phi}{\partial t} &= \sin(u). \end{aligned} \tag{15}$$

Multiplying (13) with $\cos(\phi)$ and (14) with $\sin(\phi)$ and adding, we obtain

$$\begin{aligned} 0 = & -\sin(u)\cos(\omega)\cos(\phi) - \cos(u)\sin(\omega)\frac{\partial \omega}{\partial u}\cos(\phi) \\ & -\sin(u)\sin(\omega)\sin(\phi) + \cos(u)\cos(\omega)\frac{\partial \omega}{\partial u}\sin(\phi) \end{aligned}$$

which implies that

$$\cos(u)\sin(\phi - \omega)\frac{\partial \omega}{\partial u} = \sin(u)\cos(\phi - \omega). \tag{16}$$

On the other hand, since the meridians are mapped to straight lines, we have that

$$\frac{\partial\phi}{\partial u} = 0. \tag{17}$$

Differentiating relation (15) with respect to u, we have

$$\cos(\phi-\omega)\Big(\frac{\partial\phi}{\partial u}-\frac{\partial\omega}{\partial u}\Big)\frac{\partial\phi}{\partial t}+\sin(\phi-\omega)\frac{\partial^2\phi}{\partial t\,\partial u}=\cos(u)$$

hence, from (17),

$$-\cos(\phi-\omega)\frac{\partial\omega}{\partial u}\frac{\partial\phi}{\partial t}=\cos(u). \tag{18}$$

Multiplying (16) with

$$\frac{\partial\omega}{\partial u}\frac{\partial\phi}{\partial t}$$

we have

$$\cos(u)\sin(\phi-\omega)\Big(\frac{\partial\omega}{\partial u}\Big)^2\frac{\partial\phi}{\partial t}=\sin(u)\cos(\phi-\omega)\frac{\partial\omega}{\partial u}\frac{\partial\phi}{\partial t}.$$

Hence, using (15) and (18),

$$\begin{aligned}\cos(u)\sin(u)\Big(\frac{\partial\omega}{\partial u}\Big)^2&=-\sin(u)\cos(u)\\ \Longleftrightarrow\quad \sin(u)\Big(1+\Big(\frac{\partial\omega}{\partial u}\Big)^2\Big)&=0\\ \Longleftrightarrow\quad \sin(u)&=0.\end{aligned}$$

But this last equality cannot happen in an open subset of the sphere. Thus we get a contradiction which proves Proposition 5. ☐

Acknowledgements. We would like to thank Vincent Alberge and Athanase Papadopoulos for very helpful suggestions and essential observations which improved this work significantly in many ways.

References

[1] L. Euler, "De repraesentatione superficiei sphaericae super plano," *Acta Academiae Scientarum Imperialis Petropolitinae* (1777), 107–132. Reprinted in *Opera omnia,* Series 1, Vol. 28, 248–275.
`http://eulerarchive.maa.org/pages/E490.html`
☞ 125, 126, 134

[2] L. Euler (transl. G. Heine), “On the mapping of spherical surfaces onto the plane,” tranl. of [1].
`http://eulerarchive.maa.org/docs/translations/E490en.pdf`
☞ 126

[3] L. Euler, “Principes de la trigonométrie sphérique tirés de la méthode des plus grands et des plus petits,” *Mémoires de l’Académie des sciences de Berlin* 9 (1753), 1755, 233–257. Reprinted in *Opera omnia,* Series 1, Vol. XXVII, 277–308.
`http://eulerarchive.maa.org/pages/E214.html`
☞ 127

[4] R. Gabler, J. Petersen, and L. M. Trapasso, *Essentials of physical geography,* 7th ed., Brooks/Cole, Belmont, CA, 2003.
☞ 125

Chapter 10

Area preserving maps from the sphere to the Euclidean plane

Charalampos Charitos

Contents

1. Introduction 135
2. The distortion of area-preserving geographical projections 138
3. Proof of the main theorem 141

References 149

1 Introduction

The problem of cartography, that is, the problem of constructing the mappings from the sphere S^2 (or from a subset of S^2) to the Euclidean plane E^2 which are “best” in some sense is an ancient problem, and several prominent mathematicians have been studying it from antiquity to our days. The word “best” refers to mappings which present the least distortion with respect to distances, angles or areas, or a combination of these.

In the sequel, by a mapping from S^2 to E^2 we mean a map defined either on the whole sphere S^2 or, on a subset of S^2.

In 1777, Euler published three memoirs on cartography:

(i) “De repraesentatione superficiei sphaericae super plano” (On the representation of spherical surfaces on a plane) [2];

(ii) “De proiectione geographica superficiei sphaericae” (On the geographical projections of spherical surfaces) [3];

(iii) "De proiectione geographica Deslisliana in mappa generali imperii russici usitata" (On Delisle's geographic projection used in the general map of the Russian empire) [4].

In the first memoir, he studied the following three kinds of maps:

(1) maps where the images of all meridians are perpendicular to a given horizontal axis and all parallels are parallel to it;

(2) maps which are conformal;

(3) maps which preserve area and such that the images of all meridians intersect the images of the parallels at right angles.

All these maps considered by Euler will be referred to as *geographical projections*.

Two years later, Lagrange published two memoirs on cartography [6, 7] in which he observed that there exist infinitely many, essentially different, angle-preserving (conformal) projections from S^2 to E^2 in the following sense: any two are not equal and they remain so after a composition by a similarity of E^2. Subsequently, Lagrange dealt with the problem of finding the minimal deviation map amongst all these conformal projections. To that end, he defined at each point a *magnification ratio*, which is a deviation factor from conformality at this point.

In the nineteenth century, the French mathematician and cartographer Nicolas-Auguste Tissot made a thorough investigation of geographical projections from S^2 to E^2 and he studied several of their distortion properties. Following his ideas [11], we shall recall the definition of the distortion $D_f(p)$ of f at a point $p \in S_1$, see Definition 2.2 below, where $f: S_1 \to S_2$ is a non-conformal map of class C^1 between two Riemannian surfaces. After the work of Tissot, the term *quasiconformal map* was generated and the problem of existence and uniqueness of the best quasiconformal map between Riemann surfaces was revealed as an important and difficult problem.

For the origin and development of cartography and its relation with quasiconformal maps, we refer the reader to the paper by Papadopoulos [9] which inspired the author to write the present chapter. The paper [10], also by Papadopoulos, contains biographical material on Tissot, and a review of some of his main ideas. We note incidentally that both Grötzsch and Teichmüller, who are usually referred for the early work on quasiconformal maps, refer to Tissot; cf. the history in [9] and [10].

We now explain the results.

The sphere S^2 is parameterized by the map

$$\xi\colon (t,u) \in [-\pi,\pi] \times \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \longrightarrow (\cos(t)\cos(u), \sin(t)\cos(u), \sin(u)). \tag{1}$$

For t fixed, the u-curves $\xi(t,\cdot)$ are the *meridians* of S^2 and are geodesics for the spherical metric. Setting $u = 0$, the curve $\xi(\cdot, 0)$ is the *equator* of S^2 while the t-curves $\xi(\cdot, u)$, $u \neq 0$ are circles parallel to the equator and they are called *parallels* of S^2. The parallels are not geodesics. The circle $\xi(0, \cdot)$ will be called *prime meridian.* The pair (t, u) define the geographical coordinates of S^2, namely t is the *latitude* and u is the *longitude* of S^2. Thus, the equator and the prime meridian serve as axes of S^2 with respect to which the geographical coordinates are measured.

In what follows, we always assume that ξ is restricted to the open set $(-\pi, \pi) \times \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ because on this set ξ is one-to-one. We shall denote by S_0 the subset $\xi\left((-\pi, \pi) \times \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)\right) \subset S^2$. By means of ξ, each point $e \in S_0$ with $e = \xi(t, u)$ will be identified with the pair (t, u) if no confusion is likely.

A differentiable map $f(t, u) = (x(t, u), y(t, u))$ from S_0 to E^2 is called *regular* if $\left(\frac{\partial x}{\partial t}, \frac{\partial y}{\partial t}\right) \neq (0, 0)$ and $\left(\frac{\partial x}{\partial u}, \frac{\partial y}{\partial u}\right) \neq (0, 0)$. We denote by $\mathcal{Q}$ the class of all regular maps of class C^r, $r \geq 2$ from S_0 to E^2 which preserve area and which send all meridians and parallels to perpendicular curves of E^2. In fact, the elements of $\mathcal{Q}$ are all the geographical projections which satisfy the third hypothesis in Euler's work [2] to which we referred to above. It is not hard to see that the elements of $\mathcal{Q}$ cannot be conformal maps.

If $f(t, u) = (x(t, u), y(t, u)) \in \mathcal{Q}$, we set

$$m_f(t, u) = \sqrt{\left(\frac{\partial x}{\partial u}\right)^2 + \left(\frac{\partial y}{\partial u}\right)^2}. \tag{2}$$

One can observe that $m(t, u)$ is the length in E^2 of the tangent vector to the curve $f(t, \cdot)$. We will be writing $m(t, u)$ if there is no confusion on the map f.

Finally, there is a simple area-preserving geographical projection $f_0 \in \mathcal{Q}$ which was introduced and studied by Lambert in [8, Section 6] (see also [12], pp.76–85). This projection, known as *Lambert's cylindrical equal area projection*, was rediscovered by Euler in [2]. It is defined by the formula

$$f_0(t, u) = (t, \sin(u)).$$

The goal of this note is to prove that the distortion $D_{f_0}(t, u)$ of f_0 has remarkable extremal properties with respect to the distortion of the maps $f \in \mathcal{Q}$, namely, for every t and u fixed the value $D_{f_0}(t, u)$ appears either as a maximum or as a minimum of the values $\{D_f(t, u),\ f \in \mathcal{Q}\}$.

We have the following theorem which permits us to estimate the deviation from conformality of the mappings $f \in \mathcal{Q}$, where the conformality structure on S^2 is induced from the standard Riemannian metric $ds = \sqrt{\cos^2(u) dt^2 + du^2}$ of S^2.

Theorem 1.1. *If $f(t, u) = (x(t, u), y(t, u)) \in \mathcal{Q}$, then for each $(t, u) \in S_0$,*

(i) $D_{f_0}(t, u) \leq D_f(t, u)$ *if* $m_f(t, u) \geq 1$ *or* $m_f(t, u) \leq \cos(u)$;

(ii) $D_{f_0}(t, u) \geq D_f(t, u)$ *if* $\cos(u) \leq m_f(t, u) \leq 1$.

In the last paragraph, following Euler's method, we will find explicit maps $f \in \mathcal{Q}$ and we will show that some of them satisfy the inequality $m_f(t,u) \leq \cos(u)$ for each $(t,u) \in S_0$. Therefore, the distortion of these maps is bounded below by the distortion of f_0 for each $(t,u) \in S_0$.

Remark 1.2. Tissot in [11, Chapitre troisième] studied concrete geographical projections which preserve area and which he referred to as "projections authaliques" (see [11], № 70). He made a thorough analysis of these projections and he calculated their distortion, in the sense of Definition 2.2 below, at certain points of the sphere including the equator and the poles. Using this, Tissot compiled numerical tables containing, among other things, useful information about the distortion of these projections (see the Complement in [11]). However, since all projections considered by Tissot are contained in the set $\mathcal{Q}$, Theorem 1.1 can be used to find estimates of their distortion which become quite accurate near the poles and the equator. For instance, for any map $f \in \mathcal{Q}$, the distortion $D_f(t,u)$ goes to infinity (resp. to 1) as (t,u) approaches the poles (resp. the equator) provided that $D_f(t,u)$ is bounded below (resp. above) by $D_{f_0}(t,u)$.

2 The distortion of area-preserving geographical projections

Let A be a linear isomorphism from E^2 to E^2 which is not conformal and let us denote by $\|\cdot\|$ the Euclidean norm. Let C_r be the circle of radius $r > 0$ centered at the origin O. It is well known that each C_r is sent by A to an ellipse L_r centered at O. Furthermore, there is a unique pair v_1, v_2 of perpendicular unit vectors of E^2 such that Av_1, Av_2 are perpendicular. This result was proven by Tissot in ([11], Chapitre Premier, № 3) by a simple continuity argument. Below, we reprove it in a geometric way.

The following proposition summarizes basic and nice properties of A which are used below.

Proposition 2.1. (1) *All the ellipses L_r are homothetic with homothety center O. If a_r is the great axis of L_r and b_r its small axis, then $a_1/b_1 = a_r/b_r$ for each r.*

The ratio a_1/b_1 is called the ***distortion*** *of A.*

(2) *There is a unique pair v_1, v_2 of perpendicular unit vectors of E^2 such that Av_1, Av_2 are perpendicular. Thus, if $\|v_1\| = \|v_2\|$ and if $\|Av_1\| > \|Av_2\|$ then $a_1/b_1 = \|Av_1\|/\|Av_2\|$.*

(3) *A and A^{-1} have the same distortion.*

Proof. Let a_1, b_1 be the great and small axis of L_1 respectively and let w_1, w_2 be normal vectors in the direction of the great and small axis of L_1 respectively with $\|w_1\| = a_1$, $\|w_2\| = b_1$. Let v_1, v_2 be unit vectors such that $A(v_1) = w_1$ and $A(v_2) = w_2$. Obviously $\|w_1\| < \|A(v)\| < \|w_2\|$ for each unit vector $v \neq v_1, v_2$. From the linearity of A we have also $\|A(\lambda v_1)\| < \|A(\lambda v)\| < \|A(\lambda v_2)\|$ and $\|A(\lambda v_1)\|/\|A(\lambda v_2)\| = \|w_1\|/\|w_2\| = a_1/b_1$. Thus statement (1) follows easily.

In order to show that the vectors v_1 and v_2 are perpendicular, we consider a circle C'_a (resp. C'_b) centered at the origin and tangent to L_1 at the two points where the great axis (resp. small axis) intersects L_1. Obviously C'_b is contained in the interior of L_1 and L_1 is contained in the interior of C'_a. Let K_a (resp. K_b) be the image of C'_a (resp. C'_b) by A^{-1}. Then K_a, K_b are homothetic ellipses centered at the origin which must be tangent to C_1 at the endpoints of the vectors v_1, v_2. This implies immediately that v_1, v_2 are perpendicular and unique and thus statement (2) is proven. Finally statement (3) follows easily from (2). ☐

The above properties of linear isomorphisms of E^2 give rise to the following definition of the distortion of a diffeomorphism of class C^1 between Riemannian surfaces, i.e. surfaces equipped with Riemannian metrics.

Definition 2.2 (distorsion). Let S_1, S_2 be two Riemannian surfaces and let $f: S_1 \to S_2$ be a diffeomorphism of class C^1. At each point $p \in S_1$ the differential df_p between the tangent spaces T_pS_1 of S_1 and the tangent space $T_{f(p)}S^2$ sends the unit circle C of T_pS_1 centered at the origin to an ellipse $L = df_p(C)$. Denoting by a and b the great and the small axis respectively of L, the *distortion* $D_f(p)$ of f at p is the distortion of the linear map df_p and is defined as the ratio a/b.

In Definition 2.2, if df_p preserves the angles then $D_f(p) = 1$ since $a = b$. Thus, if f is conformal then $D_f(p) = 1$ for each $p \in S_1$.

The real function $D_f: S_1 \to \mathbb{R}$ will be referred to as the *distortion function* of f.

As a special case, let $f: \Omega \to f(\Omega)$ be a C^1-diffeomorphism between the plane domains Ω and $f(\Omega)$. Let $z = (t, u) = t + iu$ be a coordinate system on Ω and let $w = f(z) = x + iy$ be a coordinate system on $f(\Omega)$. If we consider on Ω and $f(\Omega)$ the standard Euclidean metric, the distortion of f at $z_0 \in \Omega$ may be defined as previously and will be denoted by $\bar{D}_f(z_0)$, as a reminder of the fact that it is defined with respect to the Euclidean metric.

The following well-known formula can be found in many books (see for example [5], p. 2)

$$\bar{D}_f(z_0) = \frac{|f_z(z_0)| + |f_{\bar{z}}(z_0)|}{|f_z(z_0)| - |f_{\bar{z}}(z_0)|} \geq 1. \tag{3}$$

where $f_z = 1/2(f_t - if_u)$ and $f_{\bar{z}} = 1/2(f_t + if_u)$.

Obviously, if S_1, S_2 are Riemann surfaces and f is a C^1-diffeomorphism between them, the distortion $\bar{D}_f(p)$ can be defined for each $p \in S_1$. This can be done by covering S_1 (resp. S_2) by an atlas of charts which defines the Riemann structure on S_1 (resp. on S_2).

First, we will calculate the distortion function of the map

$$\xi\colon (-\pi,\pi)\times\Big(-\frac{\pi}{2},\frac{\pi}{2}\Big)\subset E^2 \longrightarrow S_0,$$

given by formula (1). The differential $d\xi_{(t,u)}$ of the map ξ sends the vectors $v_1 = (1,0)$ and $v_2 = (0,1)$ of the tangent space $T_{(t,u)}E^2$ to the vectors

$$r_1 = (-\sin(t)\cos(u), \cos(t)\cos(u), 0),$$

$$r_2 = (-\cos(t)\sin(u), -\sin(t)\sin(u), \cos(u))$$

respectively, of the tangent space $T_{\xi(t,u)}S^2 \subset E^3$. If we denote by $\langle\cdot,\cdot\rangle$ the Euclidean dot product, we have

$$\langle r_1, r_2\rangle = 0, \quad \|r_1\| = \cos(u), \quad \|r_2\| = 1.$$

This implies that the unit circle C in $T_{(t,u)}E^2$ is sent to an ellipse $L = d\xi_{(t,u)}(C)$ in $T_{\xi(t,u)}S^2$ whose great axis has length 1 and the small one has length $\cos(u)$. Therefore, the distortion $D_\xi(t,u)$ of the map ξ at (t,u) is $1/\cos(u)$. Thus, if we consider the inverse of the map ξ, say $\eta = \xi^{-1}\colon S_0 \to (-\pi,\pi)\times(-\frac{\pi}{2},\frac{\pi}{2})$, and if $e = \xi(t,u)$, we have

$$D_\eta(e) = \frac{1}{\cos(u)} \tag{4}$$

since from Proposition 2.1 the distortion of $d\xi_{(t,u)}$ is equal to the distortion of $d\eta_e$, where $e = \xi(t,u)$.

Now let $U = (-\pi,\pi)\times(-\frac{\pi}{2},\frac{\pi}{2})$ and $g\colon U \to f(U) \subset E^2$ be C^1-diffeomorphism such that the differential $dg_{(t,u)}$ sends the unit vectors $v_1 = (1,0)$, $v_2 = (0,1)$ of $T_{(t,u)}U$ to perpendicular vectors, say w_1, w_2 in $T_{g(t,u)}E^2$. Set $G = g\circ\eta\colon S_0 \to \mathbb{E}^2$.

We have the following lemma.

Lemma 2.3. *Let $e = \xi(t,u) \in S_0$. The distortion $D_G(t,u)$ of G at e is given by the product*

$$D_G(t,u) = \frac{1}{\cos(u)}\bar{D}_g(t,u). \tag{5}$$

Proof. Since $d\eta_e$ sends the vectors r_1, r_2 of T_eS^2 to v_1, $v_2 \in T_{(t,u)}U$ and $dg_{(t,u)}$ sends v_1, v_2 to the perpendicular vectors w_1, w_2 in $T_{g(t,u)}E^2$, the proof follows easily combining relation (4) and statement (2) of Proposition 2.1. ☐

Lemma 2.4. *Let* $f_0\colon S_0 \to E^2$ *be the projection* $f_0(t,u) = (t, \sin(u))$. *Then*

$$D_{f_0}(t,u) = \frac{1}{\cos^2(u)}. \tag{6}$$

Proof. The differential $df_{0_{(t,u)}}$ is expressed by the matrix $A = \left(\begin{smallmatrix}1 & 0\\ 0 & \cos(u)\end{smallmatrix}\right)$. Since $A(1,0) = (1,0)$ and $A(0,1) = (0, \cos u)$, we have from statement (2) of Proposition 2.1 that $\overline{D}_{f_0}(t,u) = 1/\cos(u)$ and from Lemma 5 that $D_{f_0}(t,u) = 1/\cos^2(u)$. ☐

3 Proof of the main theorem

Let $f\colon S_0 \to E^2$, $f(t,u) = (x(t,u), y(t,u))$ be a geographical projection belonging to the set $\mathcal{Q}$. We recall that f is a regular, area-preserving map of class C^r, $r \geq 2$, which sends the families of meridians and parallels to families of perpendicular curves in E^2.

We start with the general formulas

$$dx = p du + q dt, \quad dy = r du + s dt, \tag{7}$$

where

$$p = \frac{\partial x}{\partial u}, \quad q = \frac{\partial x}{\partial t}, \quad r = \frac{\partial y}{\partial u}, \quad s = \frac{\partial y}{\partial t}.$$

Since the images of meridians and parallels are cut perpendicularly, $pq + rs = 0$ (see [1] or [2]). Therefore, there is a function $n(t,u)$ such that

$$s = -np, \quad q = nr. \tag{8}$$

Since f is regular, $n(t,u) \neq 0$ for each (t,u). Therefore, we may assume that n is positive. In fact, if n is negative we may work similarly.

From (7) and (8) we have

$$dx = p du + nr dt, \quad dy = r du - np dt. \tag{9}$$

Following an idea of Euler, we consider the points $e = (t,u)$, $e_1 = (t + dt, u)$, $e_2 = (t, u + du)$ in S_0 and we set

$$E = f(t,u), \quad E_1 = f(t + dt, u), \quad E_2 = f(t, u + du).$$

We have, as in [1],

$$\lim_{du\to 0}\frac{|e_2-e|}{|du|}=1,$$

$$\lim_{dt\to 0}\frac{|e_1-e|}{|dt|}=\cos(u),$$

$$\lim_{du\to 0}\frac{\|E_2-E\|}{|du|}=\sqrt{\left(\frac{\partial x}{\partial u}\right)^2+\left(\frac{\partial y}{\partial u}\right)^2}=\sqrt{p^2+r^2},$$

$$\lim_{dt\to 0}\frac{\|E_1-E\|}{|dt|}=\sqrt{\left(\frac{\partial x}{\partial t}\right)^2+\left(\frac{\partial y}{\partial t}\right)^2}=n\sqrt{p^2+r^2}.$$

Therefore, the projection f is area-preserving if and only if

$$n(p^2+r^2)|du||dt|=\cos(u)|du||dt|$$

or

$$n=\frac{\cos(u)}{p^2+r^2}. \tag{10}$$

From (9) and (10) we have

$$dx=pdu+\frac{r\cos(u)}{p^2+r^2}dt,\quad dy=rdu-\frac{p\cos(u)}{p^2+r^2}dt.$$

To simplify the calculations, we set

$$p=m\cos(\phi)\quad\text{and}\quad r=m\sin(\phi),$$

where

$$m=\sqrt{\left(\frac{\partial x}{\partial u}\right)^2+\left(\frac{\partial y}{\partial u}\right)^2}>0,$$

and we have

$$dx=m\cos(\phi)du+\frac{\cos(u)\sin(\phi)}{m}dt,\quad dy=m\sin(\phi)du-\frac{\cos(u)\cos(\phi)}{m}dt.$$

Setting $m(t,u) = k(t,u)\cos(u)$ we have $k(t,u) > 0$ for each (t,u) and the relations above take the form

$$dx = k\cos(u)\cos(\phi)du + \frac{\sin(\phi)}{k}dt, \quad dy = k\cos(u)\sin(\phi)du - \frac{\cos(\phi)}{k}dt. \tag{11}$$

We may verify immediately that a solution of (11) is obtained by setting $k = 1$, $\phi = \frac{\pi}{2}$. Thus we have

$$x(t,u) = t, \quad y(t,u) = \sin(u)$$

which gives Lambert's equal area projection.

Remark 3.1. Setting $k = a$, $\phi = \varphi_0$, where a, φ_0 are arbitrary constants, the formulas

$$x(t,u) = a\sin(u)\cos(\varphi_0) + \frac{t\sin(\varphi_0)}{a}, \quad y(t,u) = a\sin(u)\sin(\varphi_0) - \frac{t\cos(\varphi_0)}{a}$$

determine a more general solution of (11). This solution differs from the previous one only in that the meridians are no longer normal to the x-axis. Actually, it is not hard to show that if k is constant in the relation (11) then the angle ϕ must be constant.

Now we can prove Theorem 1.1, which is restated below.

Theorem 1.1*. *For each* $f \in \mathfrak{Q}$,

(i) *if* $m_f(t,u) \geq 1$ *or* $m_f(t,u) \leq \cos(u)$, *then*

$$D_{f_0}(t,u) = \frac{1}{\cos^2(u)} \leq D_f(t,u);$$

(ii) *if* $\cos(u) \leq m_f(t,u) \leq 1$, *then*

$$D_{f_0}(t,u) = \frac{1}{\cos^2(u)} \geq D_f(t,u).$$

Proof. From (11) we have

$$\frac{\partial x}{\partial(u)} = k\cos(u)\cos(\phi), \quad \frac{\partial x}{\partial t} = \frac{\sin(\phi)}{k},$$

$$\frac{\partial y}{\partial u} = k\cos(u)\sin(\phi), \quad \frac{\partial y}{\partial t} = -\frac{\cos(\phi)}{k}.$$

Therefore, we may compute immediately the Jacobian of f and verify that it sends the tangent vectors $v_1 = (1,0)$ and $v_2 = (0,1)$ of $T_{(t,u)}E^2$ to the perpendicular vectors $w_1 = \big(\sin\big(\frac{\phi}{k}\big), -\cos\big(\frac{\phi}{k}\big)\big)$ and $w_2 = (k\cos(u)(\phi), k\cos(u)\sin(\phi))$ of $T_{f(t,u)}\mathbb{E}^2$. Thus Lemma 2.3 can be applied and we have

$$D_f(t,u) = \frac{1}{\cos(u)}\bar{D}_f(t,u).$$

In order to compute the term $\bar{D}_f(t,u)$, we use formula (3) and the formulas

$$A(t,u) = f_z = \frac{1}{2}(f_t - if_u) = \frac{1}{2}\Big(\Big(\frac{\partial x}{\partial t} + \frac{\partial y}{\partial u}\Big) + i\Big(\frac{\partial y}{\partial t} - \frac{\partial x}{\partial u}\Big)\Big), \tag{12a}$$

$$B(t,u) = f_{\bar{z}} = \frac{1}{2}(f_t + if_u) = \frac{1}{2}\Big(\Big(\frac{\partial x}{\partial t} - \frac{\partial y}{\partial u}\Big) + i\Big(\frac{\partial y}{\partial t} + \frac{\partial x}{\partial u}\Big)\Big). \tag{12b}$$

Therefore the equality

$$f_z = \frac{1}{2}\Big(\Big(\frac{\sin(\phi)}{k} + k\cos(u)\sin(\phi)\Big) + i\Big(-\frac{\cos(\phi)}{k} - k\cos(u)\cos(\phi)\Big)\Big)$$

implies

$$\begin{aligned}|f_z| &= \frac{1}{2}\sqrt{\Big(\frac{\sin(\phi)}{k} + k\cos(u)\sin(\phi)\Big)^2 + \Big(-\frac{\cos(\phi)}{k} - k\cos(u)\cos(\phi)\Big)^2}\\ &= \frac{1}{2}\sqrt{\frac{1}{k^2} + k^2\cos^2(u) + 2\cos(u)}\\ &= \frac{1}{2}\sqrt{\Big(\frac{1}{k} + k\cos(u)\Big)^2}\\ &= \frac{1}{2}\Big|\frac{1}{k} + k\cos(u)\Big|.\end{aligned}$$

So,

$$|A(t,u)| = \frac{1}{2}\Big|\frac{1}{k(t,u)} + k(t,u)\cos(u)\Big| = \frac{1}{2}\Big(\frac{1}{k(t,u)} + k(t,u)\cos(u)\Big) > 0. \tag{13a}$$

Similarly, we may prove that

$$|B(t,u)| = \frac{1}{2}\Big|\frac{1}{k(t,u)} - k(t,u)\cos(u)\Big|. \tag{13b}$$

We also have

$$1 \le 1/\sqrt{\cos(u)} \le \frac{1}{\cos(u)}.$$

These quantities separate the real line into four intervals. Therefore, we distinguish the following cases, depending on which interval the value $k(t,u)$ is located in.

CASE I: $k(t,u) \geq 1/\sqrt{\cos(u)}$. This is equivalent to

$$m(t,u) = k(t,u)\cos(u) \geq \sqrt{\cos(u)}.$$

Then, $(k(t,u))^2 \geq 1/\cos(u) \implies 1/k(t,u) - k(t,u)\cos(u) \leq 0$ and from (13) we have

$$|B(t,u)| = \frac{1}{2}\left|\frac{1}{k(t,u)} - k(t,u)\cos(u)\right| = \frac{1}{2}\Big(-\frac{1}{k(t,u)} + k(t,u)\cos(u)\Big). \tag{14}$$

Therefore, from (13a) and (14), we have

$$\bar{D}_f(t,u) = \frac{|A(t,u)| + |B(t,u)|}{|A(t,u)| - |B(t,u)|} = (k(t,u))^2\cos(u). \tag{15}$$

Now we distinguish the following two subcases.

SUBCASE I$_\text{a}$: $k(t,u) \geq 1/\cos(u)$. We have $m(t,u) = k(t,u)\cos(u) \geq 1$, and thus

$$k(t,u) \geq \frac{1}{\cos(u)} \implies (k(t,u))^2 \geq \frac{1}{\cos^2(u)}.$$

Combining the last relation with (15),

$$\bar{D}_f(t,u) = (k(t,u))^2\cos(u) \geq \frac{1}{\cos(u)} = \bar{D}_{f_0}(t,u)$$

and Lemma 2.3 implies that

$$D_{f_0}(t,u) = \frac{1}{\cos^2(u)} \leq D_f(t,u) = (k(t,u))^2.$$

SUBCASE I$_\text{b}$: $1/\sqrt{\cos(u)} \leq k(t,u) \leq 1/\cos(u)$. We have $\sqrt{\cos(u)} \leq m(t,u) \leq 1$, and therefore the equality

$$\bar{D}_f(t,u) = (k(t,u))^2\cos(u) \leq \bar{D}_{f_0}(t,u) = \frac{1}{\cos(u)}$$

implies

$$D_f(t,u) = (k(t,u))^2 \leq D_{f_0}(t,u) = \frac{1}{\cos^2(u)}.$$

CASE II: $k(t,u) \leq 1/\sqrt{\cos(u)}$. This is equivalent to

$$m(t,u) = k(t,u)\cos(u) \leq \sqrt{\cos(u)}.$$

First, we remark that

$$\begin{aligned} k(t,u) \leq 1/\sqrt{\cos(u)} &\iff 1/k(t,u) - k(t,u)\cos(u) \geq 0 \\ \implies |B| = \frac{1}{2}\left|\frac{1}{k(t,u)} - k(t,u)\cos(u)\right| &= \frac{1}{2}\Big(\frac{1}{k(t,u)} - k(t,u)\cos(u)\Big). \end{aligned}$$

Therefore,

$$\bar{D}_f(t,u) = \frac{|A(t,u)| + |B(t,u)|}{|A(t,u)| - |B(t,u)|} = \frac{1}{(k(t,u))^2\cos(u)}. \tag{16}$$

We distinguish the following two subcases.

SUBCASE II$_\mathrm{a}$: $k(t,u) \leq 1$. We have $m(t,u) \leq \cos(u)$. So, from (16), the equality

$$\bar{D}_{f_0}(t,u) = \frac{1}{\cos(u)} \leq \bar{D}_f = \frac{1}{(k(t,u))^2\cos(u)}$$

implies

$$D_{f_0}(t,u) = \frac{1}{\cos^2(u)} \leq D_f = \frac{1}{(k(t,u))^2\cos^2 u}.$$

SUBCASE II$_\mathrm{b}$: $1 \leq k(t,u) \leq 1/\sqrt{\cos(u)}$. We have $m(t,u) \leq \sqrt{\cos(u)}$. Thus the equality

$$\bar{D}_f(t,u) = \frac{1}{k^2\cos(u)} \leq \bar{D}_{f_0}(t,u) = \frac{1}{\cos(u)}$$

implies

$$D_f(t,u) = \frac{1}{k^2\cos^2(u)} \leq D_{f_0}(t,u) = \frac{1}{\cos^2(u)}. \qquad \square$$

Remark 3.2. For a geographical projection $f \in \mathfrak{Q}$ we may compute the distortion $\bar{D}_f(t,u)$ at a point $(t,u) \in S_0$ in a more direct way, without using formula (3). Indeed, we have seen that the differential $df_{(t,u)}$ sends the tangent vectors $v_1 = (1,0)$ and $v_2 = (0,1)$ of $T_{(t,u)}E^2$ to the perpendicular vectors $w_1 = (\sin(\phi)/k, -\cos(\phi)/k)$ and $w_2 = (k\cos(u)\cos(\phi), k\cos(u)\sin(\phi))$ of $T_{f(t,u)}E^2$ respectively. On the other hand, we have $\|w_1\| = 1/k$ and $\|w_2\| = k\cos(u)$. Therefore, considering the cases I$_\mathrm{a}$, I$_\mathrm{b}$, II$_\mathrm{a}$, and II$_\mathrm{b}$ above, we may find which of the the values $\|w_1\|$ and $\|w_2\|$ are greater and calculate the term $\bar{D}_f(t,u)$.

Now, we will show that there exist geographical projections f from S_0 to E^2 which satisfy $m_f(t,u) \leq \cos(u)$ for each $(t,u) \in (-\pi,\pi) \times \left(-\frac{\pi}{2},\frac{\pi}{2}\right)$. In fact in [2] (see p. 36ff of the corresponding translation), Euler resolves the system of equations (11) finding particular solutions as follows.

Setting $v = \sin(u)$ in (11), we get

$$dx = k\cos(\phi)dv + \frac{\sin(\phi)}{k}dt, \quad dy = k\sin(\phi)dv - \frac{\cos(\phi)}{k}dt.$$

Assuming that k is a function only of v which is equal to V and the angle ϕ is a function only of t and equal to T,

$$dx = V\cos(T)dv + \frac{\sin(T)}{V}dt, \quad dy = V\sin(T)dv - \frac{\cos(T)}{V}dt. \tag{17}$$

From equations (17) we obtain

$$x = \cos(T)\int Vdv = \frac{1}{V}\int \sin(T)dt \tag{18a}$$

and

$$y = \sin(T)\int Vdv = -\frac{1}{V}\int \cos(T)dt. \tag{18b}$$

Therefore,

$$V\int Vdv = \frac{1}{\cos(T)}\int \sin(T)dt = a$$

and

$$-V\int Vdv = \frac{1}{\sin(T)}\int \cos(T)dt = -a.$$

where a is a constant. Differentiating the relation $\int Vdv = a/V$ with respect to the variable v, we have

$$Vdv = -\frac{adV}{V^2} \implies dv = -\frac{adV}{V^3}$$

and integrating we have

$$v + c = \frac{a}{2V^2} \implies V = \sqrt{\frac{a}{2(v+c)}}, \tag{19}$$

where c is a constant and $a/(v+c) > 0$.

Similarly, for the function T, we have

$$\int \sin(T)dt = a\cos(T) \quad \text{and} \quad \int \cos(T)dt = -a\sin(T).$$

Differentiating one of these relations we have

$$dT = -\frac{dt}{a} \implies T = -\frac{t}{a}.$$

Finally, for (19) by a simple integration, we obtain

$$\int V dv = \sqrt{2a(v+c)},$$

hence

$$\begin{cases} x = \sqrt{2a(v+c)}\cos\left(\frac{t}{a}\right) = \sqrt{2a(\sin(u)+c)}\cos\left(\frac{t}{a}\right), \\ y = -\sqrt{2a(v+c)}\sin\left(\frac{t}{a}\right) = \sqrt{2a(\sin(u)+c)}\sin\left(\frac{t}{a}\right). \end{cases} \tag{20}$$

From (20), it follows that

$$\begin{cases} \frac{y}{x} = -\tan\left(\frac{t}{a}\right), \\ \sqrt{x^2+y^2} = \sqrt{2a(v+c)}. \end{cases} \tag{21}$$

From the first equation in (21), it is clear that all meridians are represented by straight lines emerging as radii from a single fixed point. The second formula implies that all parallels are portrayed as concentric circles. These projections are referred to as *azimuthal geographical projections*.

Finally from (20), we have

$$m^2(t,u) = \left(\frac{\partial x}{\partial u}\right)^2 + \left(\frac{\partial y}{\partial u}\right)^2 = \frac{\cos^2(u)}{2a(\sin(u)+c)}$$

and it is not difficult to see that there are constants a and c such that $m(t,u) \leq \cos(u)$ for each $(t,u) \in S_0$.

Acknowledgements. The author would like to thank Vincent Alberge and Athanase Papadopoulos for very helpful suggestions and observations which improved this work in many ways.

References

[1] C. Charitos and I. Papadoperakis, “On the non-existence of a perfect map from the 2-sphere to the Euclidean plane,” in *Eighteen essays in non-Euclidean geometry* (V. Alberge and A. Papadopoulos, eds.), European Mathematical Society (EMS), Zürich, 2019, Chapter 9, 125–134.
☞ 141, 142

[2] L. Euler, “De repraesentatione superficiei sphaericae super plano,” *Acta Academiae Scientarum Imperialis Petropolitinae* (1777), 107–132. Reprinted in *Opera omnia,* Series 1, Vol. 28, 248–275. Tranlation by G. Heine available on the Euler Archive.
`http://eulerarchive.maa.org/pages/E490.html`
☞ 135, 137, 141, 147

[3] L. Euler, “De proiectione geographica superficiei sphaericae,” *Acta Academiae Scientarum Imperialis Petropolitinae* (1777), 133–142. Reprinted in *Opera omnia,* Series 1, Vol. 28, 276–287. Tranlation by G. Heine avaliable on the Euler Archive.
`http://eulerarchive.maa.org/pages/E491.html`
☞ 135

[4] L. Euler, “De proiectione geographica Deslisliana in mappa generali imperii russici usitata,” *Acta Academiae Scientarum Imperialis Petropolitinae* (1777), 143–153. Reprinted in *Opera omnia,* Series 1, Vol. 28, 288–297. Tranlation by G. Heine avaliable on the Euler Archive.
`http://eulerarchive.maa.org/pages/E492.html`
☞ 136

[5] A. Fletcher and V. Markovic, *Quasiconformal maps and Teichmüller theory,* Oxford Graduate Texts in Mathematics 11. Oxford University Press, Oxford, 2007.
MR 2269887 Zbl 1103.30002 ☞ 139

[6] J.-L. Lagrange, “Sur la construction des cartes géographiques, Premier mémoire,” *Nouveaux mémoires de l’Académie royale des Sciences et Belles-Lettres de Berlin* 1779, 161–185. Reprinted in *Œuvres de Lagrange*, Vol. 4, 637–664.
☞ 136

[7] J.-L. Lagrange, “Sur la construction des cartes géographiques, Second mémoire,” *Nouveaux mémoires de l’Académie royale des Sciences et Belles-Lettres de Berlin* (1779), 186–210. Reprinted in *Œuvres de Lagrange*, Vol. 4, 664–692.
☞ 136

[8] J. H. Lambert, *Beiträge zum Gebrauche der Mathematik und deren Anwendung,* Vol. III, im Verlage des Buchladens der Realschule, Berlin 1772.
☞ 137

[9] A. Papadopoulos, "Quasiconformal mappings, from Ptolemy's geography to the work of Teichmüller," in *Uniformization, Riemann–Hilbert correspondence, Calabi–Yau manifolds, and Picard–Fuchs equations* (L. Ji and S.-T. Yau, eds.), Advanced Lectures in Mathematics 42, Higher Education Press Beijing, and International Press, Boston, 2018, 237–315.
☞ 136

[10] A. Papadopoulos, "Nicolas-Auguste Tissot: a link between cartography and quasiconformal theory," *Arch. Hist. Exact Sci.* 71 (2017), no. 4, 319–336.
MR 3665710 Zbl 1381.01007 ☞ 136

[11] N.-A. Tissot, *Mémoire sur la représentation des surfaces et les projections des cartes géographiques,* Gauthier-Villars, Paris, 1881.
☞ 136, 138

[12] John P. Snyder, *Map projections,* A Working Manual, U.S. Geological Survey Professional Paper 1395, Washington, 1987.
☞ 137

Chapter 11

Area and volume in non-Euclidean geometry

Nikolay Abrosimov and Alexander Mednykh

Contents

1. Geometry of polygons 151
2. Geometry of polyhedra 167

References 183

1 Geometry of polygons

1.1 Heron's theorem

Heron of Alexandria (c. 60 BC) is credited with the following formula that relates the area S of a triangle to its side lengths a, b, c:

$$S^2 = (s-a)(s-b)(s-c)s, \tag{1}$$

where $s = \frac{1}{2}(a+b+c)$ is the semiperimeter.

In the present section of our work we deal with the hyperbolic plane instead of the Euclidean one. The hyperbolic plane under consideration is equipped with a Riemannian metric of constant curvature $k = -1$. All the necessary definitions from hyperbolic geometry can be found in the book [8].

By definition, a *cyclic polygon* in the hyperbolic plane is a convex polygon inscribed in a curve of constant geodesic curvature (i.e. circle, horocycle or one branch of an equidistant line). Useful information about cyclic polygons can be found in [62] and [63]. In particular, it is shown in [62] that any cyclic polygon in the hyperbolic plane is uniquely determined (up to isometry) by the ordered sequence of its side lengths. In addition, among all hyperbolic polygons with fixed positive side lengths there exist polygons of maximal area. Every such a maximal polygon is cyclic (see [63]).

The following four non-Euclidean versions of Heron's formula in the hyperbolic plane are known since a long time.

Theorem 1.1 (Heron's formula in $\mathbb{H}^2$). *Let T be a hyperbolic triangle with side lengths a, b, c. Then its area $S = S(T)$ is given by each of the following formulas*:

(i) *sine of $\frac{1}{2}$ area formula,*

$$\sin^2\left(\frac{1}{2}S\right) = \frac{\sinh(s-a)\sinh(s-b)\sinh(s-c)\sinh(s)}{4\cosh^2\left(\frac{1}{2}a\right)\cosh^2\left(\frac{1}{2}b\right)\cosh^2\left(\frac{1}{2}c\right)};$$

(ii) *tangent of $\frac{1}{4}$ area formula,*

$$\tan^2\left(\frac{1}{4}S\right) = \tanh\left(\frac{1}{2}(s-a)\right)\tanh\left(\frac{1}{2}(s-b)\right)\tanh\left(\frac{1}{2}(s-c)\right)\tanh\left(\frac{1}{2}s\right);$$

(iii) *sine of $\frac{1}{4}$ area formula,*

$$\sin^2\left(\frac{1}{4}S\right) = \frac{\sinh\left(\frac{1}{2}(s-a)\right)\sinh\left(\frac{1}{2}(s-b)\right)\sinh\left(\frac{1}{2}(s-c)\right)\sinh\left(\frac{1}{2}s\right)}{\cosh\left(\frac{1}{2}a\right)\cosh\left(\frac{1}{2}b\right)\cosh\left(\frac{1}{2}c\right)};$$

(iv) *Bilinski formula,*

$$\cos\left(\frac{1}{2}S\right) = \frac{\cosh(a)+\cosh(b)+\cosh(c)+1}{4\cosh\left(\frac{1}{2}a\right)\cosh\left(\frac{1}{2}b\right)\cosh\left(\frac{1}{2}c\right)}.$$

The first two formulas are contained in the book [8], p. 66. The third formula can be obtained by square root extraction of the product of the first two. The fourth one was derived by Stanko Bilinski in [14] (see also [63]).

It should be noted that the analogous formulas in spherical space were known before the hyperbolic ones. For example, the spherical version of (i) is called Cagnoli's Theorem (see [58], Section 100), (ii) is called Lhuilier's Theorem (see [58], Section 102), formula (iv) is due to Euler (see [25]). The proofs of (iii) and (iv) can be found in [58] (Section 103).

1.2 Ptolemy's theorem and annex

We recall a few well-known facts about cyclic quadrilaterals. A convex Euclidean quadrilateral with interior angles A, B, C, D is cyclic if and only if $A+C = B+D = \pi$. A similar result for a hyperbolic quadrilateral was obtained by V. F. Petrov [46] and L. Wimmer [65]. They proved the following result.

Theorem 1.2 (Petrov, 2009, [46]; Wimmer, 2011, [65]). *A convex hyperbolic quadrilateral with interior angles A, B, C, D is cyclic if and only if $A + C = B + D$.*

Note that the sum of angles of a hyperbolic quadrilateral is less than 2π. Hence, for any cyclic hyperbolic quadrilateral we have $A + C = B + D < \pi$.

It was shown in [62] that a cyclic n-gon is uniquely determined up to isometry by the lengths of its sides. Denote the side and diagonal lengths of a quadrilateral as indicated in Figure 1. Then, in the Euclidean case, a quadrilateral is cyclic if and only if $ef = a\,c + b\,d$. This is Ptolemy's theorem. A spherical version of Ptolemy's theorem was established by Darboux and Frobenius. It can be found in the book [37] (p. 180, Proposition 4). In hyperbolic geometry, Ptolemy's identity was proposed by T. Kubota [34] and J. E. Valentine [60].

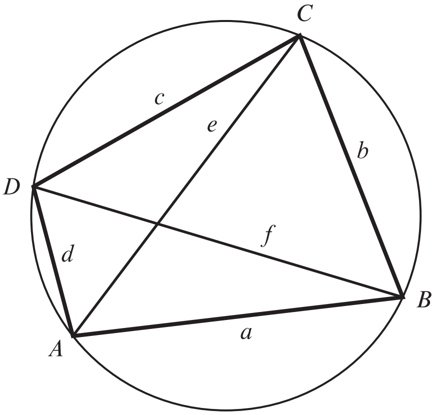


Figure 1. A cyclic quadrilateral Q with angles A, B, C, D

Theorem 1.3 (Ptolemy's identity in $\mathbb{H}^2$). *A convex hyperbolic quadrilateral with side lengths a, b, c, d and diagonal lengths e, f is cyclic if and only if*

$$\sinh\left(\frac{1}{2}e\right)\sinh\left(\frac{1}{2}f\right) = \sinh\left(\frac{1}{2}a\right)\sinh\left(\frac{1}{2}c\right) + \sinh\left(\frac{1}{2}b\right)\sinh\left(\frac{1}{2}d\right).$$

An important supplement to Ptolemy's theorem is the following property of a cyclic quadrilateral in the Euclidean plane. Its side and diagonal lengths are related by the equation

$$\frac{e}{f} = \frac{a\,d + b\,c}{a\,b + c\,d}. \tag{2}$$

Together with Ptolemy's theorem this equation allows one to express the diagonal lengths of a cyclic quadrilateral in terms of its side lengths.

It was noted in [28] that a similar relationship between the sides and diagonals of a cyclic quadrilateral is valid also in hyperbolic geometry. To make the statement true, one can replace the side length a by the quantity $s(a) = \sinh\left(\frac{1}{2}a\right)$. In particular, formula (2) can be rewritten in the following way.

Theorem 1.4. *The side lengths a, b, c, d and diagonal lengths e, f of a cyclic hyperbolic quadrilateral are related by the following equation*

$$\frac{s(e)}{s(f)} = \frac{s(a)s(d) + s(b)s(c)}{s(a)s(b) + s(c)s(d)}.$$

By making use of Theorems 1.3 and 1.4 we derive the following formulas for the diagonal lengths e, f of a cyclic hyperbolic quadrilateral. Then we have

$$s^2(e) = \frac{s(a)s(d) + s(b)s(c)}{s(a)s(b) + s(c)s(d)}(s(a)s(c) + s(b)s(d)), \tag{3}$$

$$s^2(f) = \frac{s(a)s(b) + s(c)s(d)}{s(a)s(d) + s(b)s(c)}(s(a)s(c) + s(b)s(d)). \tag{4}$$

Note that formulas (3) and (4) take place also in Euclidean and spherical geometries. In these cases, instead of the function $s(a)$ one should take the functions $s(a) = a$ and $s(a) = \sin\left(\frac{1}{2}a\right)$ respectively. See [23] and [28] for the arguments in the spherical case.

The above theorems will be used in the next section to obtain a few versions of Brahmagupta's formula for a cyclic hyperbolic quadrilateral.

1.3 Brahmagupta's formula

For polygons with more than three sides, the side lengths do not in general determine the area, but they do if the polygon is convex and cyclic (inscribed in a circle). Brahmagupta, in the seventh century, gave the following formula for a convex cyclic quadrilateral with side lengths a, b, c, d:

$$S^2 = (s-a)(s-b)(s-c)(s-d), \tag{5}$$

where $s = \frac{1}{2}(a+b+c+d)$. See [21], p. 56 for an elementary proof. An interesting consideration of the problem can be found in Möbius's paper [41]. Independently, D. P. Robbins [50] and V. V. Varfolomev [61] found a way to generalize these formulas. The main idea of both papers was to determine the squared area S^2 as a root of an algebraic equation whose coefficients are integer polynomials in the squares of the side lengths. See also [26], [18], and [51] for more detailed considerations.

One can consider Brahmagupta's formula as a generalization of Heron's formula (1).

In this section we consider the four versions of Brahmagupta's formula for a cyclic hyperbolic quadrilateral given by the second named author [38]. They are generalizations of the respective statements (i)–(iv) of Theorem 1.1.

In particular, the first statement (i) has the following analog.

Theorem 1.5 (sine of $\frac{1}{2}$ area formula). *Let Q be a cyclic hyperbolic quadrilateral with side lengths a, b, c, d. Then its area $S = S(Q)$ is given by the formula*

$$\sin^2\left(\frac{1}{2}S\right) = \frac{1}{4\mathfrak{N}}\sinh(s-a)\sinh(s-b)\sinh(s-c)\sinh(s-d)(1-\varepsilon),$$

where

$$\mathfrak{N} = \cosh^2\left(\frac{1}{2}a\right)\cosh^2\left(\frac{1}{2}b\right)\cosh^2\left(\frac{1}{2}c\right)\cosh^2\left(\frac{1}{2}d\right),$$

$$\varepsilon = \frac{\sinh\left(\frac{1}{2}a\right)\sinh\left(\frac{1}{2}b\right)\sinh\left(\frac{1}{2}c\right)\sinh\left(\frac{1}{2}d\right)}{\cosh\left(\frac{1}{2}(s-a)\right)\cosh\left(\frac{1}{2}(s-b)\right)\cosh\left(\frac{1}{2}(s-c)\right)\cosh\left(\frac{1}{2}(s-d)\right)}$$

and

$$s = \frac{1}{2}(a+b+c+d).$$

Note that the value ε vanishes if $d = 0$. In this case, we get formula (i) again.

Proofs of this theorem and subsequent theorems will be given later, at the end of this section.

The second statement for the case of a hyperbolic quadrilateral has the following form.

Theorem 1.6 (tangent of $\frac{1}{4}$ area formula). *Let Q be a cyclic hyperbolic quadrilateral with side lengths a, b, c, d. Then its area $S = S(Q)$ is given by the formula*

$$\tan^2\left(\frac{1}{4}S\right) = \frac{\sinh\left(\frac{1}{4}(-a+b+c+d)\right)\sinh\left(\frac{1}{4}(a-b+c+d)\right)}{\cosh\left(\frac{1}{4}(a+b-c-d)\right)\cosh\left(\frac{1}{4}(a-b+c-d)\right)} \frac{\sinh\left(\frac{1}{4}(a+b-c+d)\right)\sinh\left(\frac{1}{4}(a+b+c-d)\right)}{\cosh\left(\frac{1}{4}(a-b-c+d)\right)\cosh\left(\frac{1}{4}(a+b+c+d)\right)}.$$

It follows from Theorem 1.5 that for any $a, b, c, d \neq 0$ we have $1 - \varepsilon > 0$ and $\varepsilon > 0$. Hence, $0 < \varepsilon < 1$. Taking into account these inequalities as an immediate consequence of Theorems 1.5 and 1.6 we obtain the following corollary.

Corollary 1.7. *For any cyclic hyperbolic quadrilateral the following inequalities take place*:

$$\sin^2\left(\frac{1}{2}S\right) < \frac{1}{4\mathfrak{N}}\sinh(s-a)\sinh(s-b)\sinh(s-c)\sinh(s-d)$$

and

$$\tan^2\left(\frac{1}{2}S\right) > \tanh\left(\frac{1}{2}(s-a)\right)\tanh\left(\frac{1}{2}(s-b)\right)\tanh\left(\frac{1}{2}(s-c)\right)\tanh\left(\frac{1}{2}(s-d)\right).$$

By squaring the product of the formulas in Theorems 1.5 and 1.6 we obtain the following result. It can be considered as a direct generalization of statement (iii) in Theorem 1.1.

Theorem 1.8 (sine of $\frac{1}{4}$ area formula). *Let Q be a cyclic hyperbolic quadrilateral with side lengths a, b, c, d. Then its area $S = S(Q)$ is given by the formula*

$$\sin^2\left(\frac{1}{4}S\right) = \frac{1}{\mathfrak{N}'}\sinh\left(\frac{1}{2}(s-a)\right)\sinh\left(\frac{1}{2}(s-b)\right)\sinh\left(\frac{1}{2}(s-c)\right)\sinh\left(\frac{1}{2}(s-d)\right),$$

where

$$\mathfrak{N}' = \cosh\left(\frac{1}{2}a\right)\cosh\left(\frac{1}{2}b\right)\cosh\left(\frac{1}{2}c\right)\cosh\left(\frac{1}{2}d\right)$$

and

$$s = \frac{1}{2}(a+b+c+d).$$

The analogous formula in spherical space can be obtained by replacing sinh and cosh by sin and cos correspondingly (see [37], p. 182, Proposition 5).

Consider a circumscribed quadrilateral Q with side lengths a, b, c, d (Figure 2). In this case we have $a+c = b+d$, hence $s-a = c$, $s-b = d$, $s-c = a$, $s-d = b$. As a result we obtain the following assertion.

Corollary 1.9 (Brahmagupta's formula for bicentric quadrilateral). *Let Q be a bicentric* (*i.e. inscribed and circumscribed*) *hyperbolic quadrilateral with side lengths a, b, c, d. Then its area $S = S(Q)$ is given by the formula*

$$\sin^2\left(\frac{1}{4}S\right) = \tanh\left(\frac{1}{2}a\right)\tanh\left(\frac{1}{2}b\right)\tanh\left(\frac{1}{2}c\right)\tanh\left(\frac{1}{2}d\right).$$

The analogous formula in spherical space can be given by replacing tanh with tan (see [37], p. 46). A Euclidean version of this result is well known (see, e.g., [31]). In that case

$$S^2 = a\,b\,c\,d.$$

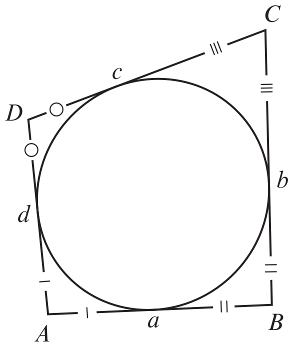


Figure 2. A circumscribed cyclic quadrilateral Q with side lengths a, b, c, d

The next theorem [38] presents a version of the Bilinski formula for a cyclic quadrilateral.

Theorem 1.10 (Bilinski formula). *Let Q be a cyclic hyperbolic quadrilateral with side lengths a, b, c, d. Then its area $S = S(Q)$ is given by the formula*

$$\cos\left(\frac{1}{2}S\right) = \frac{1}{4\mathfrak{N}'}\Big(\cosh a + \cosh b + \cosh c + \cosh d \\ - 4\sinh\left(\frac{1}{2}a\right)\sinh\left(\frac{1}{2}b\right)\sinh\left(\frac{1}{2}c\right)\sinh\left(\frac{1}{2}d\right)\Big).$$

Now we outline the proof of Theorems 1.5, 1.6, 1.8, and 1.10.

Proof. Consider a cyclic hyperbolic quadrilateral Q with side lengths a, b, c, d and interior angles A, B, C, D shown in Figure 1. By the Gauss–Bonnet formula we get the area

$$S = S(Q) = 2\pi - A - B - C - D.$$

To prove Theorem 1.5 let us find the quantities $\sin^2\left(\frac{1}{4}S\right)$ and $\cos^2\left(\frac{1}{4}S\right)$ in terms of a, b, c, d. Since $A + C = B + D$ (see Proposition 1.2) we have

$$2\sin^2\left(\frac{1}{4}S\right) = 1 - \cos\left(\frac{1}{2}S\right) = 1 - \cos(\pi - (A + C)) = 1 + \cos(A + C).$$

Hence,

$$\sin^2\left(\frac{1}{4}S\right) = \frac{1}{2}(1 + \cos(A)\cos(C) - \sin(A)\sin(C)). \tag{6}$$

Now we show that $\cos(A)$, $\cos(C)$, and the product $\sin(A)\sin(C)$ can be expressed in terms of elementary functions in a, b, c, d. To find $\cos(A)$ we use the Cosine rule for the hyperbolic triangle ABD.

$$\cos(A) = \frac{\cosh(a)\cosh(d) - \cosh(f)}{\sinh(a)\sinh(d)}. \tag{7}$$

We note that $\cosh(f) = 2\,s^2(f) + 1$, $\cosh(a) = 2\,s^2(a) + 1$, and $\cosh(d) = 2\,s^2(d) + 1$. Putting these identities into equations (4) and (7) we express $\cos(A)$ in terms of a, b, c, d. After a straightforward calculation we obtain

$$\begin{aligned}\cos(A) = \frac{1}{\widetilde{\mathfrak{N}}}(s^2(a) - s^2(b) - s^2(c) + s^2(d)\\ + 2s(a)s(b)s(c)s(d) + 2s^2(a)s^2(d)),\end{aligned} \tag{8}$$

where

$$\widetilde{\mathfrak{N}} = 2(s(a)s(d) + s(b)s(c))\cosh\left(\frac{1}{2}a\right)\cosh\left(\frac{1}{2}d\right).$$

In a similar way we get the formula

$$\begin{aligned}\cos(C) = \frac{1}{\widetilde{\mathfrak{N}}'}(-s^2(a) + s^2(b) + s^2(c) - s^2(d)\\ + 2s(a)s(b)s(c)s(d) + 2s^2(b)s^2(c)),\end{aligned} \tag{9}$$

where

$$\widetilde{\mathfrak{N}}' = 2(s(a)s(d) + s(b)s(c))\cosh\left(\frac{1}{2}b\right)\cosh\left(\frac{1}{2}c\right)$$

Note that $\sin(A)\sin(C) > 0$ and $\sin^2(A)\sin^2(C) = (1 - \cos^2(A))(1 - \cos^2(C))$. Then we take square roots in the latter equation, where $\cos(A)$ and $\cos(C)$ are given by formulas (8) and (9). Thus we express the product $\sin(A)\sin(C)$ in terms of a, b, c, d. Substituting $\cos(A)$, $\cos(C)$, and $\sin(A)\cdot\sin(C)$ into (6) and simplifying we get

$$\begin{aligned}\sin^2\left(\frac{1}{4}S\right) = \frac{1}{\mathfrak{N}'}\sinh\left(\frac{1}{4}(-a + b + c + d)\right)\sinh\left(\frac{1}{4}(a - b + c + d)\right)\\ \sinh\left(\frac{1}{4}(a + b - c + d)\right)\sinh\left(\frac{1}{4}(a + b + c - d)\right).\end{aligned} \tag{10}$$

This proves Theorem 1.8.

In a similar way, from the identity $\cos^2\left(\frac{1}{4}S\right) = 1 + \cos\left(\frac{1}{2}S\right) = 1 - \cos(A + C)$ we have

$$\begin{aligned}\cos^2\left(\frac{1}{4}S\right) = \frac{1}{\mathfrak{N}'}\cosh\left(\frac{1}{4}(a + b - c - d)\right)\cosh\left(\frac{1}{4}(a - b + c - d)\right)\\ \cosh\left(\frac{1}{4}(a - b - c + d)\right)\cosh\left(\frac{1}{4}(a + b + c + d)\right).\end{aligned} \tag{11}$$

The following lemma can be easily proved by straightforward calculations.

Lemma 1.11. *The expression*

$$H = \frac{\cosh\left(\frac{1}{4}(a+b-c-d)\right)\cosh\left(\frac{1}{4}(a-b+c-d)\right)}{\cosh\left(\frac{1}{4}(-a+b+c+d)\right)\cosh\left(\frac{1}{4}(a-b+c+d)\right)} \frac{\cosh\left(\frac{1}{4}(a-b-c+d)\right)\cosh\left(\frac{1}{4}(a+b+c+d)\right)}{\cosh\left(\frac{1}{4}(a+b-c+d)\right)\cosh\left(\frac{1}{4}(a+b+c-d)\right)}$$

can be rewritten in the form $H = 1 - \varepsilon$, *where*

$$\varepsilon = \frac{\sinh\left(\frac{1}{2}a\right)\sinh\left(\frac{1}{2}b\right)\sinh\left(\frac{1}{2}c\right)\sinh\left(\frac{1}{2}d\right)}{\cosh\left(\frac{1}{2}(s-a)\right)\cosh\left(\frac{1}{2}(s-b)\right)\cosh\left(\frac{1}{2}(s-c)\right)\cosh\left(\frac{1}{2}(s-d)\right)}$$

and

$$s = \frac{1}{2}(a+b+c+d).$$

Taking four times the products of equations (10) and (11) we have

$$\sin^2\left(\frac{1}{2}S\right) = \frac{1}{4\mathfrak{N}}\sinh\left(\frac{1}{2}(-a+b+c+d)\right)\sinh\left(\frac{1}{2}(a-b+c+d)\right) \sinh\left(\frac{1}{2}(a+b-c+d)\right)\sinh\left(\frac{1}{2}(a+b+c-d)\right)H, \tag{12}$$

where H is the same as in Lemma 1.11.

Then the statement of Theorem 1.5 follows from equation (12), Lemma 1.11 and the evident identity $s-a = \frac{1}{2}(-a+b+c+d)$. To prove Theorem 1.6 we divide (10) by (11).

Finally, the Bilinski formula (Theorem 1.10) follows from the identity $\cos\left(\frac{1}{2}S\right) = \cos^2\left(\frac{1}{4}S\right) - \sin^2\left(\frac{1}{4}S\right)$ and the above mentioned equations (10) and (11). □

1.4 Bretschneider's formula

In 1842, Carl Bretschneider [15] related the area of an arbitrary Euclidean quadrilateral to its side lengths and the sum of two opposite angles. The area S of a Euclidean quadrilateral with side lengths a, b, c, d and opposite angles A and C is given by the formula

$$S^2 = (s-a)(s-b)(s-c)(s-d) - a\,b\,c\,d\cos^2\left(\frac{1}{2}(A+C)\right),$$

where $s = \frac{1}{2}(a+b+c+d)$ is the semiperimeter (for a proof see e.g. [47], p. 89).

The statement of the theorem remains valid if one substitutes $A + C$ with the sum of another pair of opposite angles $B + D$. By making use of the identity

$$A + B + C + D = 2\pi$$

for any Euclidean quadrilateral we can rewrite the Bretschneider's theorem in the following more symmetric way

$$S^2 = (s-a)(s-b)(s-c)(s-d) - a\,b\,c\,d\,\sin^2\Big(\frac{1}{4}(A - B + C - D)\Big).$$

Precisely, this statement allows a generalization for the case of a non-Euclidean quadrilateral. In the particular case of a cyclic quadrilateral (when $A + C = B + D$) we get a Brahmagupta's formula (5). Recall that the sum of angles is not equal to 2π anymore.

The following theorem gives a hyperbolic version of Bretschneider's formula.

Theorem 1.12 (Baigonakova and Mednykh, 2012, [11]). *The area S of a hyperbolic quadrilateral with side lengths a, b, c, d and angles A, B, C, D is given by the formula*

$$\begin{aligned}
&\sin^2\Big(\frac{1}{4}S\Big)\\
&\quad= \frac{1}{\mathfrak{N}'}\sinh\Big(\frac{1}{2}(s-a)\Big)\sinh\Big(\frac{1}{2}(s-b)\Big)\sinh\Big(\frac{1}{2}(s-c)\Big)\sinh\Big(\frac{1}{2}(s-d)\Big)\\
&\qquad- \tanh\Big(\frac{1}{2}a\Big)\tanh\Big(\frac{1}{2}b\Big)\tanh\Big(\frac{1}{2}c\Big)\tanh\Big(\frac{1}{2}d\Big)\sin^2\Big(\frac{1}{4}(A - B + C - D)\Big),
\end{aligned}$$

where $s = \frac{1}{2}(a + b + c + d)$ is the semiperimeter.

Proof. Consider the *Poincaré disk model* of the hyperbolic plane. This is the unit disc $U = \{z \in \mathbb{C} : |z| < 1\}$, provided with the metric $ds = \frac{|dz|}{1-|z|^2}$. The boundary of U consists of the infinitely distant points and called *absolute*. The geodesics in this model are circle arcs and segments orthogonal to the absolute. Let z and w be arbitrary points of the disc U, and 0 be the origin. Then the hyperbolic and Euclidean distances between these points ($\rho(z, w)$ and $|z - w|$ correspondingly) are related by the formula ([13], p. 22–123)

$$|z - w| = \frac{\sinh\big(\frac{1}{2}\rho(z, w)\big)}{\cosh\big(\frac{1}{2}\rho(0, z)\big)\cosh\big(\frac{1}{2}\rho(0, w)\big)}. \tag{13}$$

Consider a hyperbolic quadrilateral $ABCD$ (Figure 3). We suppose that the vertex A coincides with the origin 0. From (13) we find the sides of the Euclidean quadrilateral $ABCD$ with respect to the hyperbolic lengths a, b, c, d, e, f. We have

$$AB = \tanh\left(\frac{1}{2}a\right), \quad BC = \frac{\sinh\left(\frac{1}{2}b\right)}{\cosh\left(\frac{1}{2}a\right)\cosh\left(\frac{1}{2}e\right)}, \tag{14a}$$

$$AD = \tanh\left(\frac{1}{2}d\right), \quad CD = \frac{\sinh\left(\frac{1}{2}c\right)}{\cosh\left(\frac{1}{2}e\right)\cosh\left(\frac{1}{2}d\right)}. \tag{14b}$$

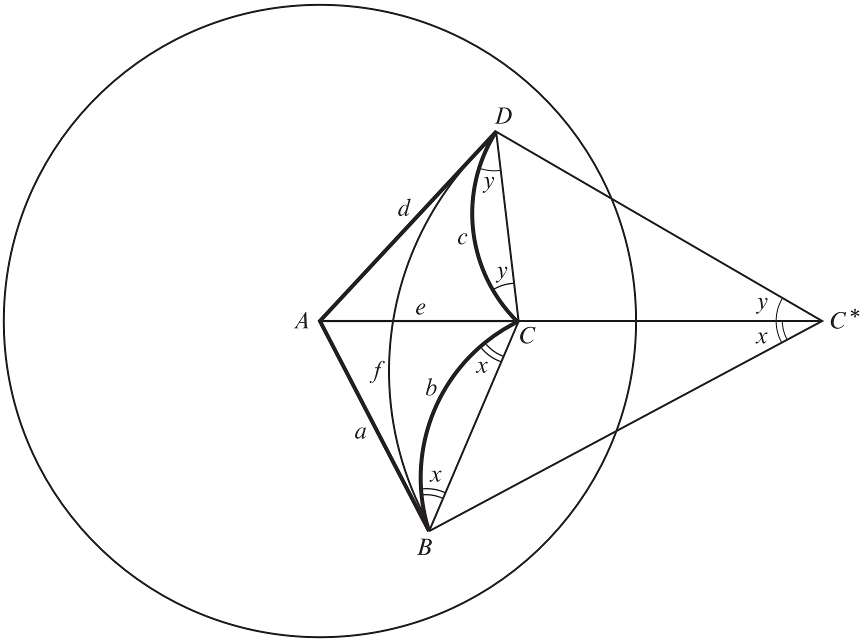


Figure 3. A hyperbolic quadrilateral $ABCD$ in the Poincaré model of $\mathbb{H}^2$

Let C^* be a point symmetric to C with respect to the absolute. To find the Euclidean distances BC^* and DC^* we use the identity $C^* = 1/C$ and the following formula from ([13], p. 123)

$$\tanh\left(\frac{1}{2}\rho(z, w)\right) = \left|\frac{z - w}{1 - z\,\overline{w}}\right|. \tag{15}$$

We get from (13) and (15)

$$BC^* = \frac{\cosh\left(\frac{1}{2}b\right)}{\cosh\left(\frac{1}{2}a\right)\sinh\left(\frac{1}{2}e\right)} \quad \text{and} \quad DC^* = \frac{\cosh\left(\frac{1}{2}c\right)}{\cosh\left(\frac{1}{2}d\right)\sinh\left(\frac{1}{2}e\right)}. \tag{16}$$

Denote the angle between the arc $\overset{\frown}{BC}$ and the chord BC by x, and the angle between the arc $\overset{\frown}{DC}$ and the chord DC by y. Then the angle measure of $\overset{\frown}{BC}$ is $2x$, while the angle measure of $\overset{\frown}{CD}$ is $2y$. Recall that the inscribed angle is equal to half of the corresponding cyclic arc. Then we obtain that the angles BC^*C and CC^*D are equal to x and y respectively. Denote by A, B, C, D the interior angles of the hyperbolic quadrilateral $ABCD$, and by $\widehat{A}$, $\widehat{B}$, $\widehat{C}$, $\widehat{D}$ the interior angles of the Eucledean quadrilateral $ABCD$. We have

$$\widehat{A} = A, \quad \widehat{B} = B + x, \quad \widehat{C} = C + x + y, \quad \widehat{D} = D + y.$$

Taking into account that $\widehat{A} + \widehat{B} + \widehat{C} + \widehat{D} = 2\pi$, we have

$$A + B + C + D + 2x + 2y = 2\pi.$$

Hence,

$$x + y = \frac{1}{2}(2\pi - A - B - C - D) = \frac{1}{2}S.$$

From the Euclidean triangle BC^*D we have

$$\sin^2\left(\frac{1}{2}(x + y)\right) = \sin^2\left(\frac{1}{4}S\right) = \frac{BD^2 - (BC^* - DC^*)^2}{4\,BC^* \cdot DC^*}.$$

Now using (14) and (16) we substitute the Euclidean lengths in the latter equation with the hyperbolic ones. Then

$$\begin{aligned}\sin^2\left(\frac{1}{4}S\right) = \frac{1}{4\mathfrak{N}'}&\left(\sinh\left(\frac{1}{2}e\right)\sinh\left(\frac{1}{2}f\right)\right)^2 \\ &- \left(\cosh\left(\frac{1}{2}a\right)\cosh\left(\frac{1}{2}c\right) - \cosh\left(\frac{1}{2}b\right)\cosh\left(\frac{1}{2}d\right)\right)^2.\end{aligned} \tag{17}$$

This formula expresses the area of a hyperbolic quadrilateral in terms of its sides and diagonals.

The next step is to find a similar formula for $\sin^2\left(\frac{1}{4}(A - B + C - D)\right)$. First, we need the following Bretschneider relation for the Euclidean quadrilateral $ABCD$ ([47], p. 85)

$$\begin{aligned}(AC \cdot BD)^2 = (AB \cdot CD)^2 + (BC \cdot AD)^2 \\ - 2AB \cdot BC \cdot CD \cdot AD\cos(\widehat{A} + \widehat{C}).\end{aligned}$$

Since

$$\widehat{A}+\widehat{C}=A+C+x+y=\pi+\frac{1}{2}(A-B+C-D),$$

we get

$$\begin{aligned}\cos\Big(\frac{1}{2}(A-B+C-D)\Big)&=-\cos(\widehat{A}+\widehat{C})\\&=\frac{(AC\cdot BD)^2-(AB\cdot CD)^2-(BC\cdot AD)^2}{2\,AB\cdot BC\cdot CD\cdot AD},\end{aligned}$$

or

$$\sin^2\Big(\frac{1}{2}(A-B+C-D)\Big)=\frac{(AB\cdot CD+BC\cdot AD)^2-(AC\cdot BD)^2}{4\,AB\cdot BC\cdot CD\cdot AD}.$$

As before, by (14) and (16) we obtain

$$\begin{aligned}&\sin^2\Big(\frac{1}{2}(A-B+C-D)\Big)\\&\quad=\frac{\big(\sinh\big(\frac{1}{2}a\big)\sinh\big(\frac{1}{2}c\big)-\sinh\big(\frac{1}{2}b\big)\sinh\big(\frac{1}{2}d\big)\big)^2-\big(\sinh\big(\frac{1}{2}e\big)\sinh\big(\frac{1}{2}f\big)\big)^2}{4\sinh\big(\frac{1}{2}a\big)\sinh\big(\frac{1}{2}b\big)\sinh\big(\frac{1}{2}c\big)\sinh\big(\frac{1}{2}d\big)}.\end{aligned}\tag{18}$$

Eliminating the expression $\big(\sinh\big(\frac{1}{2}e\big)\sinh\big(\frac{1}{2}f\big)\big)^2$ from (17) and (18) and making use of standard formulas we finally obtain

$$\begin{aligned}\sin^2\Big(\frac{1}{4}S\Big)=\frac{1}{\mathfrak{N}'}\sinh\Big(\frac{1}{2}(s-a)\Big)\sinh\Big(\frac{1}{2}(s-b)\Big)\sinh\Big(\frac{1}{2}(s-c)\Big)\sinh\Big(\frac{1}{2}(s-d)\Big)\\-\tanh\Big(\frac{1}{2}a\Big)\tanh\Big(\frac{1}{2}b\Big)\tanh\Big(\frac{1}{2}c\Big)\tanh\Big(\frac{1}{2}d\Big)\\\sin^2\Big(\frac{1}{4}(A-B+C-D)\Big).\end{aligned}$$

□

The analogous theorem holds in spherical geometry as well (for more details see the recent work [10] or the old papers [12] and [29]).

Theorem 1.13 (Baigonakova and Mednykh, 2012, [11]). *The area S of a spherical quadrilateral with side lengths a,b,c,d and angles A, B, C, D is given by the formula*

$$\begin{aligned}\sin^2\Big(\frac{1}{4}S\Big)=\frac{\sin\big(\frac{1}{2}(s-a)\big)\sin\big(\frac{1}{2}(s-b)\big)\sin\big(\frac{1}{2}(s-c)\big)\sin\big(\frac{1}{2}(s-d)\big)}{\cos\big(\frac{1}{2}a\big)\cos\big(\frac{1}{2}b\big)\cos\big(\frac{1}{2}c\big)\cos\big(\frac{1}{2}d\big)}\\-\tan\Big(\frac{1}{2}a\Big)\tan\Big(\frac{1}{2}b\Big)\tan\Big(\frac{1}{2}c\Big)\tan\Big(\frac{1}{2}d\Big)\sin^2\Big(\frac{1}{4}(A-B+C-D)\Big),\end{aligned}$$

where $s=\frac{1}{2}(a+b+c+d)$ is the semiperimeter.

1.5 Casey's theorem

John Casey generalized Ptolemy's theorem in the following way (see [16], p. 103).

Theorem 1.14 (Casey's theorem). *Let four circles O_1, O_2, O_3, O_4 on the Euclidean plane $\mathbb{E}^2$ be internally tangent to a circle O in the prescribed order. Denote by t_{ij} the length of the common external tangent of the circles O_i, O_j. Then*

$$t_{12}\,t_{34} + t_{23}\,t_{14} = t_{13}\,t_{24}\,.$$

We say that a common tangent of two circles is *external* if it does not intersect the segment joining the centers of the circles.

The following hyperbolic version of Casey's theorem is given in [7].

Theorem 1.15 (Abrosimov and Mikaiylova, 2015, [7]). *Let four circles O_1, O_2, O_3, O_4 on the hyperbolic plane $\mathbb{H}^2$ be internally tangent to a circle O in the prescribed order. Denote by t_{ij} the length of the common external tangent of the circles O_i, O_j. Then*

$$\sinh\left(\frac{1}{2}t_{12}\right)\sinh\left(\frac{1}{2}t_{34}\right) + \sinh\left(\frac{1}{2}t_{23}\right)\sinh\left(\frac{1}{2}t_{14}\right) = \sinh\left(\frac{1}{2}t_{13}2\right)\sinh\left(\frac{1}{2}t_{24}\right).$$

Proof. Denote by O_i the centers of the circles respectively (Figure 4). Let R_i be the radii of the circles O_i and K_i the points of tangency with the circle O.

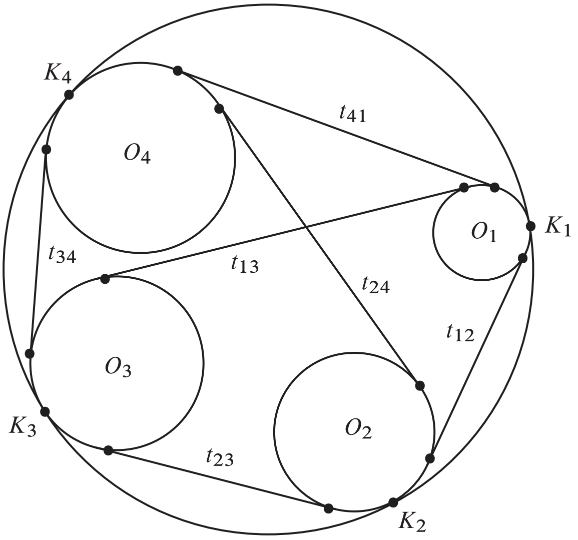


Figure 4

Consider the quadrilateral with sides $O_i O_j$, R_i, R_j, and t_{ij} (Figure 5). Denote by d the diagonal which divides the quadrilateral into triangles 1 and 2. Since the radii R_i and R_j are orthogonal to the common tangent t_{ij}, we can apply the hyperbolic Pythagorean theorem for triangle 1 (see, e.g., [8])

$$\cosh(d) = \cosh(R_j)\cosh(t_{ij}). \tag{19}$$

Denote the angle between the diagonal d and the tangent t_{ij} by δ. Then by the Sine rule for the hyperbolic triangle 1 we have

$$\sin\delta = \frac{\sinh(R_j)}{\sinh(d)}.$$

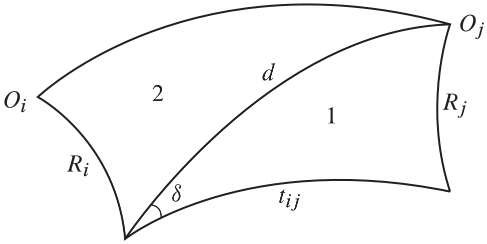


Figure 5

We use the second Cosine rule for the hyperbolic triangle 2 (see, e.g., [8])

$$\cosh(O_i O_j) = \cosh(R_i)\cosh(d) - \sinh(R_i)\sinh(R_j)\cos\left(\frac{\pi}{2} - \delta\right). \tag{20}$$

Substituting $\cosh(d)$ from (20) into (19) we obtain

$$\cosh(O_i O_j) = \cosh(t_{ij})\cosh(R_i)\cosh(R_j) - \sinh(R_i)\sinh(R_j). \tag{21}$$

We apply the first Cosine rule to the triangle OK_iK_j where $OK_i = OK_j = R$ (see Figure 6)

$$\cos(\angle K_i OK_j) = \frac{\cosh^2(R) - \cosh(K_i K_j)}{\sinh^2(R)}. \tag{22}$$

Using the first Cosine rule for the hyperbolic triangle OO_iO_j, where $OO_i = R - R_i$, $OO_j = R - R_j$, we have

$$\begin{aligned}\cosh(O_i O_j) = {} & \cosh(R - R_i)\cosh(R - R_j) \\ & - \sinh(R - R_i)\sinh(R - R_j)\cos(\angle K_i OK_j).\end{aligned}$$

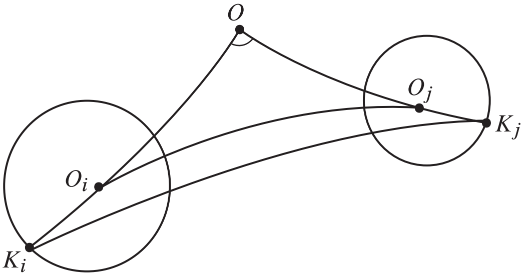


Figure 6

Combining the latter expression with (21) and (22) we obtain

$$\sinh^2\left(\frac{1}{2}K_iK_j\right) = \frac{\sinh^2(R)\cosh(R_i)\cosh(R_j)}{2\sinh(R-R_i)\sinh(R-R_j)}\sinh^2\left(\frac{1}{2}t_{ij}\right). \tag{23}$$

The points of tangency of the circles O_1, O_2, O_3, O_4 with the circle O form a cyclic quadrilateral $K_1K_2K_3K_4$ in $\mathbb{H}^2$. According to the hyperbolic Ptolemy theorem (Theorem 1.3), the segments K_iK_j satisfy the equation

$$\begin{aligned}&\sinh\left(\frac{1}{2}K_1K_3\right)\sinh\left(\frac{1}{2}K_2K_4\right)\\ &\quad=\sinh\left(\frac{1}{2}K_1K_2\right)\sinh\left(\frac{1}{2}K_3K_4\right)+\sinh\left(\frac{1}{2}K_2K_3\right)\sinh\left(\frac{1}{2}K_1K_4\right).\end{aligned} \tag{24}$$

We take the square root of (23) and substitute the expressions for $\sinh\left(\frac{1}{2}K_iK_j\right)$ in (24). Then we divide the resulting equation by

$$\sinh^2 R\sqrt{\frac{\cosh(R_1)\cosh(R_2)\cosh(R_3)\cosh(R_4)}{2\sinh(R-R_1)\sinh(R-R_2)\sinh(R-R_3)\sinh(R-R_4)}}$$

to obtain

$$\sinh\left(\frac{1}{2}t_{13}\right)\sinh\left(\frac{1}{2}t_{24}\right)=\sinh\left(\frac{1}{2}t_{12}\right)\sinh\left(\frac{1}{2}t_{34}\right)+\sinh\left(\frac{1}{2}t_{23}\right)\sinh\left(\frac{1}{2}t_{14}\right).$$

□

The spherical analog of Casey's theorem holds as well [7].

Theorem 1.16 (Abrosimov and Mikaiylova, 2015, [7]). *Let four circles O_1, O_2, O_3, O_4 on the sphere $\mathbb{S}^2$ be internally tangent to a circle O in the prescribed order. Denote by t_{ij} the length of the common external tangent of the circles O_i, O_j. Then*

$$\sin\left(\frac{1}{2}t_{12}\right)\sin\left(\frac{1}{2}t_{34}\right)+\sin\left(\frac{1}{2}t_{23}\right)\sin\left(\frac{1}{2}t_{14}\right)=\sin\left(\frac{1}{2}t_{13}\right)\sin\left(\frac{1}{2}t_{24}\right).$$

One can prove the spherical version of the theorem using the same arguments as in the hyperbolic case. The only difference is to replace hyperbolic trigonometric relations by corresponding spherical ones.

2 Geometry of polyhedra

Calculating volumes of polyhedra is a classical problem, that has been well known since Euclid. It remains relevant nowadays, partly due to the fact that the volume of a fundamental polyhedron is one of the main invariants for a three-dimensional manifold.

2.1 Volume of hyperbolic orthoscheme

Volume formulas for non-Euclidean tetrahedra in some special cases have been known since Lobachevsky, Bolyai and Schläfli. For example, Schläfli [53] found the volume of an orthoscheme in the 3-dimensional sphere $\mathbb{S}^3$. Recall that an *orthoscheme* is an n-dimensional simplex defined by a sequence of edges $(v_0, v_1), (v_1, v_2), \ldots, (v_{n-1}, v_n)$ that are mutually orthogonal. In three dimensions, an orthoscheme is also called a *birectangular tetrahedron*. By definition, a 3-dimensional orthoscheme has three dihedral angles $\frac{\pi}{2}$ and three others, which we will refer as *essential* ones (see Figure 7).

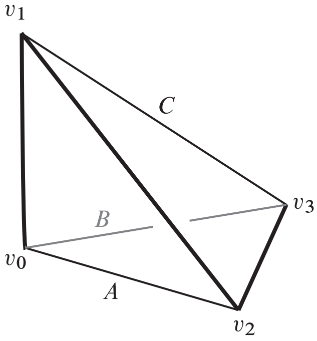


Figure 7. An orthoscheme $T = T(A, B, C)$ with essential dihedral angles A, B, C, all other angles are $\frac{\pi}{2}$

Theorem 2.1 (Schläfli, 1858, [52]). *Let T be a spherical orthoscheme with essential dihedral angles A, B, C. Then its volume $V = V(T)$ is given by the formula*

$$V = \frac{1}{4} S(A, B, C),$$

where

$$\begin{aligned} S\Big(\frac{\pi}{2} - x, y, \frac{\pi}{2} - z\Big) &= \widehat{S}(x, y, z) \\ &= \sum_{m=1}^{\infty} \Big(\frac{D - \sin(x)\sin(z)}{D + \sin(x)\sin(z)}\Big)^m \frac{\cos(2mx) - \cos(2my) + \cos(2mz) - 1}{m^2} \\ &\quad - x^2 + y^2 - z^2, \end{aligned}$$

and $D \equiv \sqrt{\cos^2(x)\cos^2(z) - \cos^2(y)}$.

The function $S(x, y, z)$ appearing in Theorem 2.1 is called the *Schläfli function*.

The volume of a hyperbolic orthoscheme was obtained independently by Nikolai Lobachevsky [35] and János Bolyai (see [57], [56], or [64]). The following theorem was suggested by G. S. M. Coxeter [20]. It represents a result of Lobachevsky in a very simple form.

Theorem 2.2 (Lobachevsky, 1835, [35]; Coxeter, 1835, [20]). *Let T be a hyperbolic orthoscheme with dihedral angles A, B, C. Then its volume $V = V(T)$ is given by the formula $V = \frac{i}{4} S(A, B, C)$, where $S(A, B, C)$ is the Schläfli function.*

J. Bolyai found the volume of a hyperbolic orthoscheme in terms of planar angles and an edge length. This result was not published during his lifetime. It can be found in the work [57], pp. 111–118 collected on the base of original manuscripts in reliquias by János Bolyai (see also [56] and [64]).

Theorem 2.3 (Bolyai, 1832). *Consider a hyperbolic orthoscheme T with essential dihedral angles along AC, CB, BD, and all other dihedral angles equal to $\frac{\pi}{2}$ (Figure 8). T is given by its planar angles α, β, γ and edge length z. Then its volume $V = V(T)$ can be found by the formula*

$$V = \frac{\tan(\gamma)}{2\tan(\beta)} \int_0^z \frac{u \sinh(u) du}{\Big(\frac{\cosh^2(u)}{\cos^2(\alpha)} - 1\Big)\sqrt{\frac{\cosh^2(u)}{\cos^2(\gamma)} - 1}}.$$

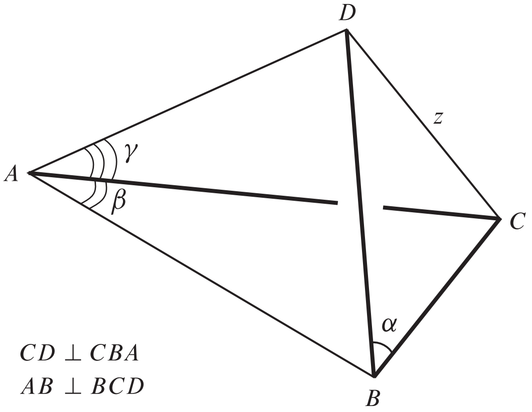


Figure 8. An orthoscheme T given by planar angles α, β, γ and edge length z

2.2 Volume of ideal tetrahedra and octahedra

An *ideal tetrahedron* is a hyperbolic tetrahedron with all vertices at infinity. Opposite dihedral angles of an ideal tetrahedron are pairwise equal and the sum of dihedral angles at the edges adjacent to one vertex is $A + B + C = \pi$ (see Figure 9).

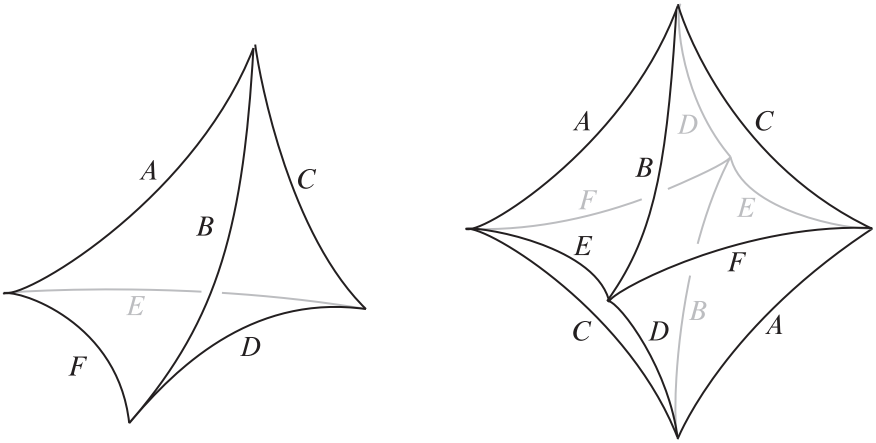


Figure 9. An ideal tetrahedron $T = T(A, B, C)$ and an ideal octahedron $O = O(A, B, C, D, E, F)$

The volume of an ideal tetrahedron has been known since Lobachevsky [35]. J. Milnor presented it in very elegant form [39].

Theorem 2.4 (Lobachevsky, 1835, [35]; Milnor, 1982, [39]). *Let T be an ideal tetrahedron with dihedral angles A, B and C. Then its volume $V = V(T)$ is given by the formula*

$$V = \Lambda(A) + \Lambda(B) + \Lambda(C),$$

where

$$\Lambda(x) = -\int\limits_0^x \log|2\sin t|dt$$

is the Lobachevsky function.

Corollary 2.5. *The maximal volume of an ideal tetrahedron is attained in the case when the dihedral angles $A = B = C = \frac{1}{3}\pi$. It is equal to*

$$V_{\max} = 3\Lambda\left(\frac{\pi}{3}\right) = 1.01494\ldots.$$

A more general case of a tetrahedron with at least one vertex at infinity was investigated by E. B. Vinberg [8].

An *ideal octahedron* is a hyperbolic octahedron with all vertices at infinity. Consider a symmetric ideal octahedron whose opposite dihedral angles are pairwise equal (see Figure 9). By definition, the sum of the dihedral angles at the edges adjacent to any vertex is equal to 2π. We have

$$A + B + C + D = 2\pi, \quad A + C + E + F = 2\pi, \quad B + D + E + F = 2\pi,$$

or

$$C = \pi - A, \quad D = \pi - B, \quad F = \pi - E.$$

Hence, $A, B, C, D, E, F \in (0, \pi)$. The volume formula for such an octahedron is very similar to Milnor's formula for ideal tetrahedra and is given in [9].

Theorem 2.6 (Baigonakova and Mednykh, 2010, [9]). *Let O be a symmetric ideal octahedron with dihedral angles A, B, C, D, E, F. Then its volume $V = V(O)$ is given by the formula*

$$\frac{\mathrm{Vol}(O)}{2} = \Lambda\left(\frac{1}{2}(\pi + A + B + E)\right) + \Lambda\left(\frac{1}{2}(\pi - A - B + E)\right) + \Lambda\left(\frac{1}{2}(\pi + A - B - E)\right) + \Lambda\left(\frac{1}{2}(\pi - A + B - E)\right),$$

where $\Lambda(x)$ is the Lobachevsky function.

Corollary 2.7. *The maximal volume of an ideal octahedron is attained in the case when $A = B = C = D = E = F = \frac{\pi}{2}$. It is equal to*

$$V_{\max} = 8\Lambda\left(\frac{\pi}{4}\right) = 3.66386\ldots.$$

2.3 Seidel's conjecture

In 1986, J. J. Seidel [54] conjectured that the volume of an ideal hyperbolic tetrahedron can be expressed as a function of the determinant and the permanent of its Gram matrix. Recall that the formula expressing the volume of such a tetrahedron in terms of dihedral angles has been known since Lobachevsky and Bolyai (Theorem 2.4). In spite of this, Seidel's problem had not been solved for a long time. Ten years later, a strengthened version of Seidel's conjecture was suggested by Igor Rivin and Feng Luo. They supposed that the volume of a non-Euclidean tetrahedron (hyperbolic or spherical one) depends only on the determinant of its Gram matrix. It was shown in [1] and [2] that the strengthened conjecture is false, while Seidel's conjecture is true within certain conditions.

Consider a non-Euclidean tetrahedron T in the sphere $\mathbb{S}^3$ or in the hyperbolic space $\mathbb{H}^3$. Denote its vertices by numbers $1, 2, 3, 4$. Let A_{ij} denote the dihedral angle at the edge joining vertices i and j. For convenience, we set $A_{ii} = \pi$ for $i = 1, 2, 3, 4$. It is well known [8] that, in the hyperbolic and spherical spaces, the tetrahedron T is uniquely (up to isometry) determined by its Gram matrix $G = \langle -\cos(A_{ij})\rangle_{i,j=1,\dots,4}$.

Recall that the permanent of a matrix $M = \langle m_{ij}\rangle_{ij=1,2,\dots,n}$ is defined by

$$\operatorname{per}(M) = \sum_{i=1}^{n} m_{ij} \operatorname{per}(M_{ij}), \quad \operatorname{per}(m_{ij}) = m_{ij},$$

where M_{ij} is the matrix obtained from M by removing the i-th row and j-th column. Conditions for the existence of spherical and hyperbolic tetrahedra in terms of Gram matrices are given in [36] and [59], respectively.

The solution of Seidel's problem, which was posed in [54], is given by the following theorem.

Theorem 2.8 (Abrosimov, 2010, [2]). *In each of the classes of acute or obtuse*[1] *ideal tetrahedra the volume is determined uniquely by the determinant and permanent of the Gram matrix.*

Proof. As is known (see, e.g., [39]), opposite dihedral angles of an ideal tetrahedron are pairwise equal and the sum of the dihedral angles at the edges adjacent to one vertex is $A + B + C = \pi$ (Figure 9).

---

[1] By an *obtuse tetrahedron* we mean a tetrahedron with at least one dihedral angle $> \frac{\pi}{2}$.

Thus, we can take $C = \pi - A - B$. The Gram matrix has the form

$$G = \begin{pmatrix} 1 & -\cos(A) & -\cos(B) & \cos(A+B) \\ -\cos(A) & 1 & \cos(A+B) & -\cos(B) \\ -\cos(B) & \cos(A+B) & 1 & -\cos(A) \\ \cos(A+B) & -\cos(B) & -\cos(A) & 1 \end{pmatrix}.$$

We have

$$\det(G) = -4\sin^2(A)\sin^2(B)\sin^2(A+B),$$

$$\operatorname{per}(G) = 4 + 4\cos^2(A)\cos^2(B)\cos^2(A+B).$$

To prove Theorem 2.8, we show that the dihedral angles of an ideal tetrahedron are uniquely (up to a permutation) determined by $\det G$ and per G in each of the cases: acute angled tetrahedron and obtuse angled tetrahedron.

Without loss of generality, we can assume that $0 < A \le B \le C = \pi - A - B$. Then the dihedral angles A, B are a priori acute, and the angle C is either acute or obtuse. In the former case, the tetrahedron under consideration is acute-angled, and in the latter case, it is obtuse-angled.

Let us introduce the new variables

$$x = \sin(A)\sin(B), \quad y = \cos(A)\cos(B)$$

and show that, for a fixed left-hand side, the solutions of the system of equations

$$-\frac{1}{4}\det(G) = x^2(1-(y-x)^2),$$

$$\frac{1}{4}\operatorname{per}(G) = y^2(y-x)^2 + 1$$

correspond to one tetrahedron (up to isometry) in each of the two cases mentioned above.

Suppose that the system has a pair of different solutions (a, b) and (x, y). Then we have

$$a^2(1-(b-a)^2) = x^2(1-(y-x)^2),$$

$$b^2(b-a)^2 = y^2(y-x)^2.$$

For the angle C to be acute means that

$$\cos(A)\cos(B) - \sin(A)\sin(B) = -\cos(C) < 0,$$

i.e., both solutions satisfy the inequalities $b(b-a) < 0$ and $x(x-y) < 0$. If the angle C is obtuse, then the reverse inequality holds. This allows us to take a square root in the second equation without loosing solutions:

$$a^2(1-(b-a)^2) = x^2(1-(y-x)^2),$$

$$b\,(b-a) = y\,(y-x).$$

Expressing x from the second equation and substituting the resulting expression into the first one, we obtain a sixth-degree polynomial equation in y. Fortunately, it decomposes into the linear factors and the biquadratic polynomial

$$(b-y)\cdot(b+y)\cdot\big(y^4 - (a^2 + a^4 + 2\,ab - 2a^3b - b^2 + 2ab^3 - b^4)y^2 \\ + a^4b^2 - 4a^3b^3 + 6a^2b^4 - 4ab^5 + b^6\big) = 0.$$

Thus, all solutions can be found in radicals. Substituting the expressions for x and y in terms of dihedral angles, we see that different solutions of the system correspond to the same ideal tetrahedron $T(A, B, C)$ up to reordering the dihedral angles. ☐

Note that, in Theorem 2.8, the assumption that the tetrahedron is either acute-angled or obtuse-angled cannot be dispensed with. This is demonstrated by the following example.

Example 2.9. Consider a pair of ideal tetrahedra, the obtuse-angled tetrahedron $T_1(s, s, \pi - 2s)$ and the acute-angled tetrahedron $T_2\big(t, \frac{1}{2}(\pi - t), \frac{1}{2}(\pi - t)\big)$, where

$$s = \arccos\left(\frac{\sqrt{2 + \sqrt{4 + \sqrt{170\sqrt{17} - 698}}}}{2\sqrt{2}}\right),$$

$$t = \arccos\left(\frac{-1 + \sqrt{17} + \sqrt{-26 + 10\sqrt{17}}}{8}\right).$$

The determinants and the permanents of the Gram matrices of these tetrahedra coincide. We have

$$\det(G(T_1)) = \det(G(T_2)) = \frac{107 - 51\sqrt{17}}{128},$$

$$\operatorname{per}(G(T_1)) = \operatorname{per}(G(T_2)) = \frac{163 + 85\sqrt{17}}{128}.$$

At the same time, the volumes of the tetrahedra T_1 and T_2 are different and equal to 0.847365 and 1.01483 respectively.

2.4 Volume of a general hyperbolic tetrahedron

Despite the fact that partial results on the volume of non-Euclidean tetrahedra were known, a volume formula for hyperbolic tetrahedra of general type remained unknown until recently. A general algorithm for obtaining such a formula was indicated by W.-Y. Hsiang in [30]. A complete solution was obtained more than ten years later. A general formula was proposed in paper by Y. Cho and H. Kim [17]. However, it was asymmetric with respect to permutation of angles. J. Murakami and Y. Yano [44] proposed a formula expressing the volume by dihedral angles in a symmetric way. A year later, A. Ushijima [59] presented a simple proof of the Murakami–Yano formula. He also investigated the case of a truncated hyperbolic tetrahedron.

It should be noted that in all these studies the volume is expressed as a linear combination of 16 dilogarithms or Lobachevsky functions. The arguments of these functions depend on the dihedral angles of the tetrahedron and some additional parameter, which is the root of some quadratic equation with complex coefficients in a sophisticated form.

The geometric meaning of that formula was explained by P. Doyle and G. Leibon [24] in terms of Regge symmetry. A clear description of these ideas and a complete geometric proof of the formula was given by Y. Mohanty [42]. In particular, she proved the equivalence of Regge symmetry and homogeneity (scissors congruence) in hyperbolic space [43].

In 2005, D. A. Derevnin and A. D. Mednykh [22] proposed the following integral formula for the volume of a hyperbolic tetrahedron.

Theorem 2.10 (Derevnin and Mednykh, 2005, [22]). *Let* $T(A, B, C, D, E, F)$ *be a compact hyperbolic tetrahedron with dihedral angles* A, B, C, D, E, F*. Then its volume* $V = V(T)$ *is given by the formula*

$$V = -\frac{1}{4}\int\limits_{z_1}^{z_2} \log \frac{\cos\left(\frac{1}{2}(A+B+C+z)\right)\cos\left(\frac{1}{2}(A+E+F+z)\right)}{\sin\left(\frac{1}{2}(A+B+D+E+z)\right)\sin\left(\frac{1}{2}(A+C+D+F+z)\right)} \frac{\cos\left(\frac{1}{2}(B+D+F+z)\right)\cos\left(\frac{1}{2}(C+D+E+z)\right)}{\sin\left(\frac{1}{2}(B+C+E+F+z)\right)\sin\left(\frac{1}{2}z\right)}\,dz,$$

where z_1 *and* z_2 *are the roots of the integrand satisfied the condition* $0 < z_2 - z_1 \le \pi$.

More precisely,

$$z_1 = \arctan\left(\frac{k_3}{k_4}\right) - \arctan\left(\frac{k_1}{k_2}\right), \quad z_2 = \pi - \arctan\left(\frac{k_3}{k_4}\right) - \arctan\left(\frac{k_1}{k_2}\right),$$

where

$$\begin{aligned} k_1 &= -\cos(S) - \cos(A+D) - \cos(B+E) - \cos(C+F) - \cos(D+E+F) \\ &\quad - \cos(D+B+C) - \cos(A+E+C) - \cos(A+B+F), \\ k_2 &= \sin(S) + \sin(A+D) + \sin(B+E) + \sin(C+F) + \sin(D+E+F) \\ &\quad + \sin(D+B+C) + \sin(A+E+C) + \sin(A+B+F), \\ k_3 &= 2\,(\sin A \sin D + \sin B \sin E + \sin C \sin F), \\ k_4 &= \sqrt{k_1^2 + k_2^2 - k_3^2}, \\ S &= A + B + C + D + E + F. \end{aligned}$$

2.5 Sforza's formula for non-Euclidean tetrahedra

Surprisingly, but more than 100 years ago, in 1907, Giuseppe Sforza found a fairly simple formula for the volume of a non-Euclidean tetrahedron. This fact has became widely known after a discussion between the second named author and J. M. Montesinos at the conference in El Burgo de Osma (Spain), August, 2006. Unfortunately, the outstanding work of Sforza [55] published in Italian has been completely forgotten.

The original arguments by Sforza are based on some identity given by H. W. Richmond [49] besides the Schläfli formula. Then he used the Pascal equation [45] for minors of the Gram matrix and some routine calculations. In this subsection we present another simpler proof of Sforza's formula. The idea of the proof belongs to the authors [6].

Consider a hyperbolic (or a spherical) tetrahedron T with dihedral angles A, B, C, D, E, F. Assume that A, B, C are dihedral angles at the edges adjacent to one vertex and D, E, F are opposite to them correspondingly (see Figure 10). Then the Gram matrix $G(T)$ is defined as follows:

$$G = \begin{pmatrix} 1 & -\cos(A) & -\cos(B) & -\cos(F) \\ -\cos(A) & 1 & -\cos(C) & -\cos(E) \\ -\cos(B) & -\cos(C) & 1 & -\cos(D) \\ -\cos(F) & -\cos(E) & -\cos(D) & 1 \end{pmatrix}.$$

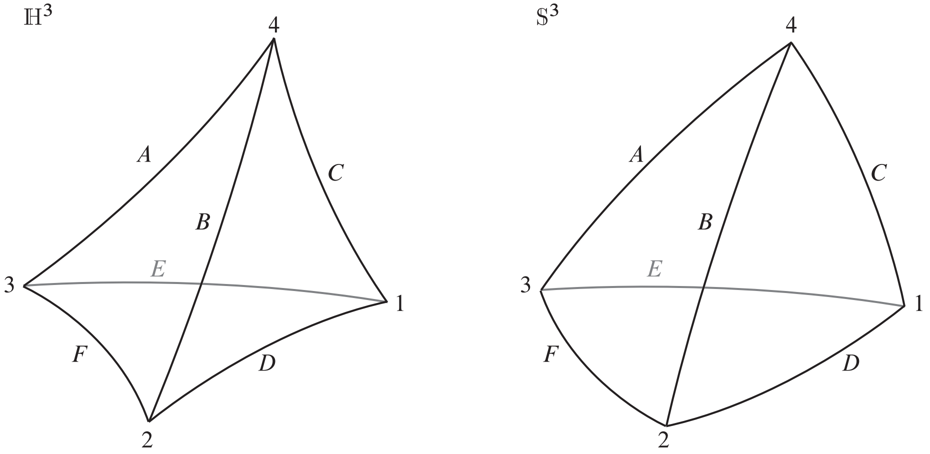


Figure 10. A tetrahedron $T = T(A, B, C, D, E, F)$ in $\mathbb{H}^3$ or $\mathbb{S}^3$

Denote by $\mathcal{C} = \langle c_{ij} \rangle_{i,j=1,2,3,4}$ a matrix of cofactors $c_{ij} = (-1)^{i+j} G_{ij}$, where G_{ij} is (i, j)-th minor of G.

In the following proposition we collect some known results about hyperbolic tetrahedra (see, e.g., [59]).

Proposition 2.11. *Let T be a compact hyperbolic tetrahedron. Then*

(i) $\det(G) < 0$;

(ii) $c_{ii} > 0,\ i \in \{1, 2, 3, 4\}$;

(iii) $c_{ij} > 0,\ i \neq j;\ i, j \in \{1, 2, 3, 4\}$;

(iv) $\cosh(\ell_{ij}) = \frac{c_{ij}}{\sqrt{c_{ii} c_{jj}}}$,

where ℓ_{ij} is the hyperbolic length of the edge joining vertices i and j.

Further, we need the following assertion due to Jacobi ([48], Theorem 2.5.1, p. 12).

Proposition 2.12 (Jacobi theorem). *Let $M = \langle m_{ij} \rangle_{i,j=1,\dots,n}$ be a matrix and $\Delta = \det(M)$ be the determinant of M. Denote by $\mathcal{C} = \langle c_{ij} \rangle_{i,j=1,\dots,n}$ the matrix of cofactors $c_{ij} = (-1)^{i+j} \det M_{ij}$, where M_{ij} is the $(n-1) \times (n-1)$ minor obtained by removing the i-th row and j-th column of the matrix M. Then for any k, $1 \leq k \leq n-1$ we have*

$$\det\langle c_{ij} \rangle_{i,j=1,\dots,k} = \Delta^{k-1} \det\langle m_{ij} \rangle_{i,j=k+1,\dots,n}.$$

One of the main tools for volume calculations in $\mathbb{H}^3$ and $\mathbb{S}^3$ is the classical Schläfli formula.

Proposition 2.13 (Schläfli's formula). *Let X^n be a space of constant curvature K. Consider a family of convex polyhedra P in X^n depending on one or more parameters in a differential manner and keeping the same combinatorial type. Then the differential of the volume $V = V(P)$ satisfies the equation*

$$(n-1)K\,dV = \sum_F V_{n-2}(F)\,d\theta(F),$$

where the sum is taken over all $(n-2)$-facets of P, $V_{n-2}(F)$ is $(n-2)$-dimensional volume of F, and $\theta(F)$ is the interior angle along F.

In the classical paper by Schläfli [52] this formula was proved for the case of a spherical n-simplex. For the hyperbolic case, it was obtained by H. Kneser [33] (for more details, see also [8] and [40]).

Now we are able to prove the following theorem (see also [6]).

Theorem 2.14 (Sforza's formula in $\mathbb{H}^3$). *Let T be a compact hyperbolic tetrahedron with Gram matrix G. Consider $G = G(A)$ as a function of the dihedral angle A. Then the volume $V = V(T)$ is given by the formula*

$$V = \frac{1}{4}\int\limits_{A_0}^{A} \log\left(\frac{c_{34}(A) - \sqrt{-\det\, G(A)}\sin(A)}{c_{34}(A) + \sqrt{-\det\, G(A)}\sin(A)}\right) dA,$$

where A_0 is a suitable root of the equation $\det\, G(A) = 0$ *and $c_{34} = c_{34}(A)$ is the* $(3,4)$*-cofactor of the matrix $G(A)$.*

Proof. Let $\Delta = \det(G)$. By the Jacobi theorem (Proposition 2.12) applied to the Gram matrix G for $n = 4$ and $k = 2$ we obtain

$$c_{33}c_{44} - c_{34}^2 = \Delta(1 - \cos^2(A)).$$

By the Cosine rule (Proposition 2.11, (iv)) we get $\cosh(\ell_A) = \frac{c_{34}}{\sqrt{c_{33}c_{44}}}$. Hence,

$$\sinh(\ell_A) = \sqrt{\cosh^2(\ell_A) - 1} = \sqrt{\frac{c_{34}^2 - c_{33}c_{44}}{c_{33}c_{44}}} = \frac{\sqrt{-\Delta}\sin(A)}{\sqrt{c_{33}c_{44}}}.$$

Since $\exp(\pm\ell_A) = \cosh(\ell_A) \pm \sinh(\ell_A)$ we have

$$\exp(\pm\ell_A) = \frac{c_{34} \pm \sqrt{-\Delta}\,\sin(A)}{\sqrt{c_{33}c_{44}}}.$$

Hence,

$$\exp(2\,\ell_A) = \frac{\exp(\ell_A)}{\exp(-\ell_A)} = \frac{c_{34} + \sqrt{-\Delta}\,\sin(A)}{c_{34} - \sqrt{-\Delta}\,\sin(A)}$$

and

$$\ell_A = \frac{1}{2}\log\Big(\frac{c_{34} + \sqrt{-\Delta}\,\sin(A)}{c_{34} - \sqrt{-\Delta}\,\sin(A)}\Big).$$

By the Schläfli formula (Proposition 2.13) for $V = V(T)$ we have

$$-dV = \frac{1}{2}\sum_{\alpha}\ell_\alpha\,d\alpha, \quad \alpha \in \{A, B, C, D, E, F\}.$$

By the assumption that the angle A is variable and all the other angles are constant, we get

$$-dV = \frac{1}{2}\,\ell_A\,dA.$$

Note that $\det(G) \to 0$ as $A \to A_0$. Thus, $V \to 0$ as $A \to A_0$. Then integrating both sides of the equation we obtain

$$V = \int\limits_{A_0}^{A}\Big(-\frac{\ell_A}{2}\Big)\,dA = \frac{1}{4}\int\limits_{A_0}^{A}\log\Big(\frac{c_{34} - \sqrt{-\det G}\,\sin(A)}{c_{34} + \sqrt{-\det G}\,\sin(A)}\Big)\,dA,$$

where the lower limit A_0 is a suitable root of the equation $\det G(A) = 0$. ☐

In the following proposition we collect some known results about spherical tetrahedra (see, e.g., [36]).

Proposition 2.15. *Let T be a spherical tetrahedron. Then*

(i) $\det(G) > 0$;

(ii) $c_{ii} > 0$, $i = 1, \dots, 4$;

(iii) $\cos(\ell_{ij}) = \frac{c_{ij}}{\sqrt{c_{ii}c_{jj}}}$,

where ℓ_{ij} is the length of the edge joining vertices i and j in $\mathbb{S}^3$.

The next theorem presents a spherical version of Sforza's formula.

Theorem 2.16 (Sforza's formula in $\mathbb{S}^3$). *Let T be a spherical tetrahedron with Gram matrix G. Consider $G = G(A)$ as a function of the dihedral angle A. Then the volume $V = V(T)$ is given by the formula*

$$V = \frac{1}{4i}\int\limits_{A_0}^{A} \log\left(\frac{c_{34}(A) + i\sqrt{\det\, G(A)}\,\sin(A)}{c_{34}(A) - i\sqrt{\det\, G(A)}\,\sin(A)}\right) dA,$$

where A_0 is a suitable root of the equation $\det(G(A)) = 0$ *and $c_{34} = c_{34}(A)$ is the* $(3,4)$*-cofactor of the matrix $G(A)$.*

The proof is similar to one given in the hyperbolic case.

2.6 Volumes of hyperbolic octahedra with symmetries

In this subsection, we consider octahedra in $\mathbb{H}^3$ and $\mathbb{S}^3$. It is known that a regular ideal octahedron is a fundamental polyhedron for the complement of the Whitehead link embedded into the three-dimensional sphere. Moreover, a hyperbolic manifold of minimum volume (the Matveev–Fomenko–Weeks manifold) and many other manifolds can be obtained by Dehn surgeries on the Whitehead link. Thus, hyperbolic octahedra can serve as fundamental polyhedra for a wide class of manifolds, including ones of small volume. The latter seems to be of particular interest when one deals with hyperbolic manifolds ordered by increasing volume.

The problem of finding a volume formula for an arbitrary non-Euclidean polyhedron is very hard. It was only solved for a general hyperbolic or spherical tetrahedron. Even for an octahedron the problem is still open. But in the case of a polyhedron with nontrivial symmetry, the calculation of the volume is simplified. This fact was first shown by Lobachevsky, who found the volume of an ideal hyperbolic tetrahedron. It follows from the definition of an ideal tetrahedron that its opposite dihedral angles are pairwise equal (i.e. it has a symmetry).

Exact formulas for the volumes of spherical octahedra that have mmm- or $2|m$-symmetry were derived in [3]. The volumes of hyperbolic octahedra with mmm- or $\bar{3}$-symmetry were calculated in [4] and [5] respectively. These studies were initiated by Galiulin, Mikhalev, and Sabitov's paper [27], in which the authors calculated the volumes of Euclidean octahedra with nontrivial symmetry. We show just two from the above mentioned results.

An octahedron is said to *have* $2|m$*-symmetry* if it admits a rotation through an angle of $\frac{2}{3}\pi$ and a mirror symmetry about a plane orthogonal to the axis of rotation (see Figure 11). The $2|m$-symmetry of the octahedron corresponds to the group C_{2h} in the Schoenflies classification, to the group $[2, 2^+]$ in the Coxeter classification, or to the group 2^* in the orbifold classification (see, e.g., [19], [32]). The volume formula for spherical octahedra with $2|m$-symmetry is given in [3].

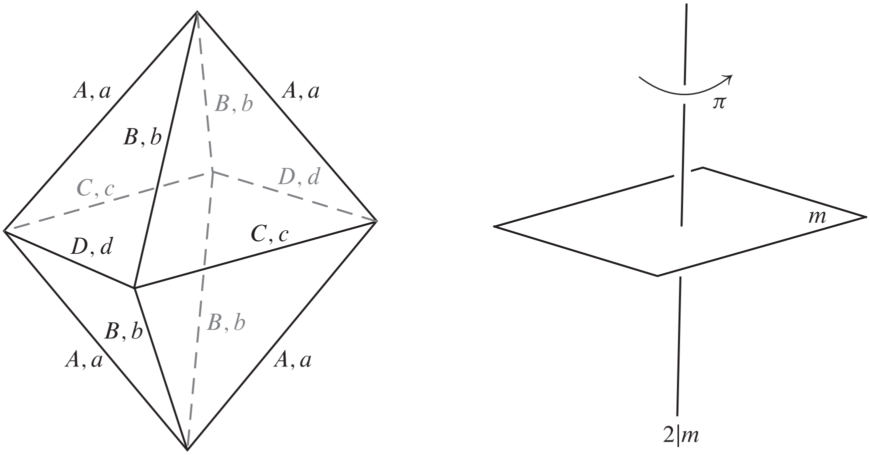


Figure 11. An octahedron $\mathcal{O} = \mathcal{O}(A, B, C, D)$ with $2|m$-symmetry

Theorem 2.17 (Abrosimov, Godoy-Molina, and Mednykh, 2008, [3]). *Let* $\mathcal{O} = \mathcal{O}(A, B, C, D)$ *be a spherical octahedron with dihedral angles* A, B, C, D *that admits a* $2|m$*-symmetry. Then its volume* $V = V(\mathcal{O})$ *is given by the formula*

$$V = 2\int\limits_{\frac{\pi}{2}}^{\theta} \Big(\operatorname{arctanh}(X\cos(\tau)) + \operatorname{arctanh}(Y\cos(\tau)) + \operatorname{arctanh}(Z\cos(\tau)) + \operatorname{arctanh}(W\cos(\tau))\Big)\frac{d\tau}{\cos(\tau)},$$

where $X = \cos(A)$*,* $Y = \cos(B)$*,* $Z = \cos\left(\frac{1}{2}(C + D)\right)$*,* $W = \cos\left(\frac{1}{2}(C - D)\right)$*, and* θ $\left(0 \le \theta \le \frac{\pi}{2}\right)$ *is a root of the equation*

$$\cos^2(\theta) + \frac{X + Y + Z + W}{XYZ + XYW + XZW + YZW} = 0.$$

Moreover, θ *is determined by the sine–tangent rule*

$$\frac{\sin(A)}{\tan(a)} = \frac{\sin(B)}{\tan(b)} = \frac{\sin\left(\frac{1}{2}(C + D)\right)}{\tan\left(\frac{1}{2}(c + d)\right)} = \frac{\sin\left(\frac{1}{2}(C - D)\right)}{\tan\left(\frac{1}{2}(c - d)\right)} = \tan(\theta).$$

An octahedron is said to *have* $\bar{3}$*-symmetry* if it admits the central symmetry and a rotation through an angle of $\frac{2}{3}\pi$ (see Figure 12). The $\bar{3}$-symmetry of the octahedron corresponds to the group S_6 in the Schoenflies classification, to the group $[2^+, 6^+]$ in the Coxeter classification, or to the group $3\times$ in the orbifold classification (see, e.g., [19] or [32], Chapter 11). The existence criterion and the volume formula for a hyperbolic octahedron with this type of symmetry are given in [5].

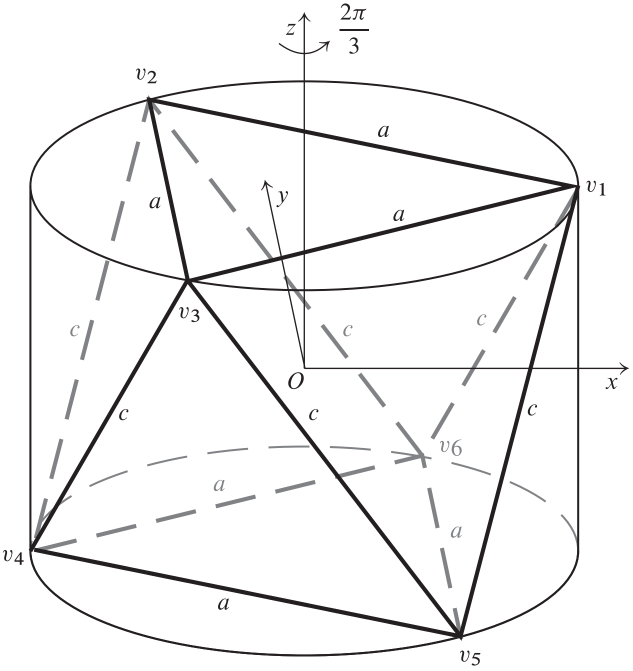


Figure 12. An octahedron $\mathcal{O} = \mathcal{O}(a, c)$ with $\bar{3}$-symmetry

Theorem 2.18 (Abrosimov, Kudina, and Mednykh, 2015, [5]). *Let* $a, c > 0$ *be some numbers. Then the following statements are equivalent:*

(i) *there exists a compact hyperbolic octahedron* $\mathcal{O}(a, c)$ *with edge lengths* a, c *that admits a* $\bar{3}$*-symmetry*;

(ii) *the inequality*

$$\frac{\sinh^2\left(\frac{1}{2}a\right)}{\sinh^2\left(\frac{1}{2}c\right)} < 3$$

holds.

Theorem 2.19 (Abrosimov, Kudina, and Mednykh, 2015, [5]). *Let* $\mathcal{O} = \mathcal{O}(a, c)$ *be a compact hyperbolic octahedron with edge lengths* a, c *that admits a* $\bar{3}$*-symmetry. Then its volume* $V = V(\mathcal{O})$ *is given by the formula*

$$V = 3 \int_0^a \frac{t\,F + (1 + 2\cosh(t))cG}{(1 + 2\cosh(t))\Delta}\,dt,$$

where

$$\begin{aligned} F = {} & 1 + 2\cosh(t) + 2\cosh^2(t) + \cosh^3(t) \\ & - 2\cosh(c) - 2\cosh(t)\cosh(c) - \cosh^2(c) \\ & - \cosh(t)\cosh^2(c) + \cosh^2(t)\cosh(c) \\ & - 4\cosh^2(t)\cosh^2(c) + 3\cosh^3(c), \end{aligned}$$

$$G = \sinh(t)\sinh(c)(-1 + 2\cosh(t)),$$

$$\Delta = (\cosh(2c) - \cosh(t))\sqrt{(-2\cosh(t) + 3\cosh(c))(\cosh(t) + \cosh(c))}.$$

The algorithm developed in [5] for obtaining the volume formula is rather universal. It can be also applied for other hyperbolic polyhedra with non-trivial symmetry.

The idea is the following. First, the polyhedron under consideration is modeled in Euclidean space. The coordinates of the vertices in $\mathbb{E}^3$ are calculated in some parameters according to the symmetry. Then the Euclidean polyhedron is placed into the projective Caley–Klein model of hyperbolic space. This gives rise to some additional conditions on the coordinates. Vector algebra in projective space is used to express hyperbolic edge lengths and dihedral angles in terms of coordinate parameters. Then symmetry is used again to exclude non-invariant parameters depending on the choice of coordinates. As a result we obtain a direct relation between lengths and angles of the hyperbolic polyhedron. These relations allow us to establish an existence criterion for such polyhedra in $\mathbb{H}^3$. Moreover, they allow us to find the solution of the Schläfli differential equation responsible for the volume. The latter provides an exact integral formula for the volume of the hyperbolic polyhedron.

Acknowledgement. The research of this article was funded by Russian Science Foundation (project no. 16-41-02006).

References

[1] N. V. Abrosimov, “Solving the Seidel problem on the volume of hyperbolic tetrahedron,” *Sib. Èlektron. Mat. Izv.* 6 (2009), 211–218, in Russian.
MR 2586687 Zbl 1299.51007 ☞ 171

[2] N. V. Abrosimov, “Seidel’s problem on the volume of a non-Euclidean tetrahedron,” *Dokl. Akad. Nauk* 435 (2010), no. 1, 7–10, in Russian, English translation in *Dokl. Math.* 82 (2010), no. 3, 843–846.
MR 2790493 Zbl 1228.52007 ☞ 171

[3] N. V. Abrosimov, M. Godoy-Molina, and A. D. Mednykh, “On the volume of a spherical octahedron with symmetries,” *Sovrem. Mat. Prilozh.* 60, Algebra (2008), 3–12, in Russian; English translation in *J. Math. Sci. (N.Y.)* 161 (2009), no. 1, 1–10.
MR 2676254 Zbl 1187.51021 ☞ 179, 180

[4] N. V. Abrosimov and G. A. Baigonakova, “Hyperbolic octahedron with mmm-symmetry,” *Sib. Èlektron. Mat. Izv.* 10 (2013), 123–140, in Russian.
MR 3039999 Zbl 1330.52014 ☞ 179

[5] N. V. Abrosimov, E. S. Kudina, and A. D. Mednykh, “On the volume of a hyperbolic octahedron with $\bar{3} - symmetry$,” *Tr. Mat. Inst. Steklova* 288 (2015), 7–15, in Russian; English translation in *Proc. Steklov Inst. Math.* 288 (2015), no. 1, 1–9.
MR 3485697 Zbl 1341.52022 ☞ 179, 181, 182

[6] N. V. Abrosimov and A. D. Mednykh, “Volumes of polytopes in constant curvature spaces,” in *Rigidity and symmetry* (R. Connelly, A. I. Weiss and W. Whiteley, eds.), Fields Institute Communications, 70, Springer, New York, and Fields Institute for Research in Mathematical Sciences, Toronto, ON, 2014, 1–26.
MR 3329266 Zbl 1320.52013 ☞ 175, 177

[7] N. V. Abrosimov and L. A. Mikaiylova, “Casey’s theorem in hyperbolic geometry,” *Sib. Èlektron. Mat. Izv.* 12 (2015), 354–360.
MR 3493779 Zbl 06607128 ☞ 164, 166

[8] D. V. Alekseevskij, E. B. Vinberg, and A. S. Solodovnikov, “Geometry of spaces of constant curvature,” in *Geometry* II (a translation of Геометрия II, Akad. Nauk SSSR, Vsesoyuz. Inst. Nauchn. i Tekhn. Inform., Moscow, 1988), translation by V. Minachin, translation edited by E. B. Vinberg, Encyclopaedia of Mathematical Sciences, 29, Springer-Verlag, Berlin, 1993, 1–138.
MR 1254932 ☞ 151, 152, 165, 170, 171, 177

[9] G. A. Baigonakova and A. D. Mednykh, “On Milnor formula for hyperbolic octahedron,” *Mat. Zamet. YAGU* 17 (2010), no. 2, 3–9.
Zbl 1274.51016 ☞ 170

[10] G. A. Baigonakova and A. D. Mednykh, “On Bretschnider formula for spherical quadrilateral,” *Mat. Zamet. YAGU* 19 (2012), no. 1, 3–11, in Russian.
Zbl 1274.51022 ☞ 163

[11] G. A. Baigonakova and A. D. Mednykh, “On Bretschnider formula for hyperbolic quadrilateral,” *Mat. Zamet. YAGU* 19 (2012), no. 1, 12–19, in Russian.
Zbl 1274.51023 ☞ 160, 163

[12] C. W. Baur, “Das Sehnenviereck in der Ebene und auf der Kugel als besonderer Fall des allgemeinen,” *Zeitschrift für Mathematik und Physik* 6 (1861), 221–234.
☞ 163

[13] A. F. Beardon, *The geometry of discrete groups,* Graduate Texts in Mathematics, 91. Springer-Verlag, New York, 1983.
MR 0698777 Zbl 0528.30001 ☞ 160, 161

[14] S. Bilinski, “Zur Begründung der elementaren Inhaltslehre in der hyperbolischen Ebene,” *Math. Ann.* 180 (1969), 256–268.
Zbl 0159.21702 ☞ 152

[15] C. A. Bretschneider, “Untersuchung der trigonometrischen Relationen des geradlinigen Viereckes,” *Archiv der Mathematik und Physik* 2, (1842), 225–261.
☞ 159

[16] J. Casey, *A sequel to the first six books of the Elements of Euclid,* containing an easy introduction to modern geometry, with numerous examples, 5th ed., Hodges, Figgis and Co., Dublin, 1888.
☞ 164

[17] Yu. Cho and H. Kim, “On the volume formula for hyperbolic tetrahedra,” *Discrete Comput. Geom.* 22 (1999), no. 3, 347–366.
MR 1706606 Zbl 0952.51013 ☞ 174

[18] R. Connelly, “Comments on generalized Heron polynomials and Robbins’ conjectures,” *Discrete Math.* 309 (2009), no. 12, 4192–4196.
MR 2537417 Zbl 1181.51021 ☞ 154

[19] J. H. Conway, H. Burgiel, and Ch. Goodman-Strass, *The symmetries of things,* A. K. Peters, Wellesley, MA, 2008.
MR 2410150 Zbl 1173.00001 ☞ 180, 181

[20] H. S. M. Coxeter, "The functions of Schläfli and Lobatschefsky," *Quart. J. Math. Oxford.* 6 (1935), no. 1, 13–29.
JFM 61.0395.02 Zbl 0011.17006 ☞ 168

[21] H. S. M. Coxeter and S. L. Greitzer, *Geometry revisited,* New Mathematical Library, 19. Random House, New York, 1967.
MR 3155265 Zbl 0166.16402 ☞ 154

[22] D. A. Derevnin and A. D. Mednykh, "A formula for the volume of a hyperbolic tetrahedron," *Uspekhi Mat. Nauk* 60 (2005), no. 2, 159–160, in Russian; English translation in *Russian Math. Surveys* 60 (2005), no. 2, 346–348.
MR 2152953 Zbl 1085.51503 ☞ 174

[23] W. Dickinson and M. Salmassi, "The right right triangle on the sphere," *College Math. J.* 39 (2008), no. 1, 24–33.
MR 2378203 Zbl 1293.97034 ☞ 154

[24] P. Doyle and G. Leibon, "23040 symmetries of hyperbolic tetrahedra," preprint, 2003.
arXiv:math/0309187v1 [math.GT]
☞ 174

[25] L. Euler, "Variae speculationes super area triangulorum sphaericorum," *Nova Acta Academiae Scientarum Imperialis Petropolitinae* 10 (1797), 47–62. Reprinted in *Opera omnia,* Series 1, Vol. 29, 253–266.
`http://eulerarchive.maa.org/pages/E698.html`
☞ 152

[26] M. Fedorchuk and I. Pak, "Rigidity and polynomial invariants of convex polytopes," *Duke Math. J.* 129 (2005), no. 2, 371–404.
MR 2165546 Zbl 1081.52012 ☞ 154

[27] R. V. Galiulin, S. N. Mikhalev, and I. Kh. Sabitov, "Some applications of the formula for the volume of an octahedron," *Mat. Zametki* 76 (2004), no. 1, 27–43, in Russian; English translation in *Math. Notes* 76 (2004), no. 1-2, 25–40.
MR 2099840 Zbl 1075.52006 ☞ 179

[28] R. Guo and N. Sönmez, "Cyclic polygons in classical geometry," *C. R. Acad. Bulgare Sci.* 64 (2011), no. 2, 185–194.
MR 2865551 Zbl 1289.51013 ☞ 154

[29] G. Hessenberg, "Elementare Beweise für eine Maximumeigenschaft des Sehnenvierecks," in *Schwarz-Festschrift,* Julius Springer, Berlin, 1914, 76–83.
☞ 163

[30] W.-Y. Hsiang, “On infinitesimal symmetrization and volume formula for spherical or hyperbolic tetrahedrons,” *Quart. J. Math. Oxford Ser.* (2) 39 (1988), no. 156, 463–468.
MR 0975910 Zbl 0664.51011 ☞ 174

[31] V. F. Ivanoff, C. F. Pinzka, and J. Lipman, “Elementary problems and solutions: solutions: E1376,” *Amer. Math. Monthly* 67 (1960), no. 3, 291–292.
MR 1530669 ☞ 156

[32] N. W. Johnson, *Geometries and transformations* Cambridge University Press, Cambridge, 2018.
Zbl 06706287 ☞ 180, 181

[33] H. Kneser, “Der Simplexinhalt in der nichteuklidischen Geometrie,” *Deutsche Math.* 1, (1936), 337–340.
JFM 62.0653.03 Zbl 0014.36203 ☞ 177

[34] T. Kubota, “On the extended Ptolemy’s theorem in hyperbolic geometry,” *Science reports of the Tohoku University,* Ser. 1: Physics, chemistry, astronomy, 2 (1912), 131–156.
☞ 153

[35] N. I. Lobatschefskij (N. I. Lobachevsky), *Imaginäre Geometrie und ihre Anwendung auf einige Integrale,* German translation by H. Liebmann, Teubner, Leipzig, 1904. ☞ 168, 169, 170

[36] F. Luo, “On a problem of Fenchel,” *Geom. Dedicata* 64 (1997), no. 3, 277–282.
MR 1440561 Zbl 0873.51015 ☞ 171, 178

[37] W. J. M’Clelland and T. Preston, *A treatise on spherical trigonometry, with application to spherical geometry and numerous examples,* Part II. Macmillian and Co., London, 1886.
☞ 153, 156

[38] A. D. Mednykh, “Brahmagupta formula for cyclic quadrilaterials in the hyperbolic plane,” *Sib. Èlektron. Mat. Izv.* 9 (2012), 247–255.
MR 2954694 Zbl 1330.51009 ☞ 155, 157

[39] J. Milnor, “Hyperbolic geometry: the first 150 years,” *Bull. Amer. Math. Soc. (N.S.)* 6 (1982), no. 1, 9–24.
MR 0634431 Zbl 0486.01006 ☞ 169, 170, 171

[40] J. Milnor, “How to compute volume in hyperbolic space,” in *Collected papers,* 1. Geometry, Publish or Perish, Houston, TX, 1994, 189–212.
☞ 177

[41] A. F. Möbius, “Ueber die Gleichungen, mittelst welcher aus den Seiten eines in einen Kreis zu beschreibenden Vielecks der Halbmesser des Kreises und die Fläche des Vielecks gefunden werden,” *Crelle’s Journal* 35 (1828), 5–34.
☞ 154

[42] Y. Mohanty, *Hyperbolic polyhedra*: *volume and scissors congruence,* Ph.D. Thesis. University of California, San Diego, 2002.
MR 2703698 ☞ 174

[43] Y. Mohanty, “The Regge symmetry is a scissors congruence in hyperbolic space,” *Algebr. Geom. Topol.* 3 (2003), 1–31.
MR 1997312 Zbl 1026.51015 ☞ 174

[44] J. Murakami and M. Yano, “On the volume of a hyperbolic and spherical tetrahedron,” *Comm. Anal. Geom.* 13 (2005), no. 2, 379–400.
MR 2154824 Zbl 1081.51012 ☞ 174

[45] E. Pascal, *I determinanti*: *teoria ed applicazioni,* U. Hoepli, Milano, 1897.
☞ 175

[46] F. V. Petrov, “Inscribed quadrilaterals and trapezoids in absolute geometry,” *Mat. Prosveshchenie* 13 (2009), 149–154, in Russian.
☞ 152, 153

[47] Ya. P. Ponarin, *Elementary planimetria,* Vol. 1, MCCME, Moscow, 2004, in Russian.
☞ 159, 162

[48] V. V. Prasolov, *Problems and theorems in linear algebra,* translated from the Russian manuscript by D. A. Leĭtes, Translations of Mathematical Monographs 134. American Mathematical Society, Providence, R.I., 1994.
MR 1277174 Zbl 0803.15001 ☞ 176

[49] H. W. Richmond, “The volume of a tetrahedron in elliptic space,” *Quart. J. Math.* 34 (1903), 175–177.
JFM 34.0679.04 ☞ 175

[50] D. P. Robbins, “Areas of polygons inscribed in a circle,” *Discrete Comput. Geom.* 12 (1994), no. 2, 223–236.
MR 1283889 Zbl 0806.52008 ☞ 154

[51] I. Kh. Sabitov, “Solution of cyclic polyhedra,” *Mat. Prosveshchenie* 14 (2010), 83–106, in Russian.
☞ 154

[52] L. Schläfli, “On the multiple integral $\int\int\cdots\int dxdy\ldots dz$ whose limits are $p_1 = a_1x+b_1y+\cdots+h_1z > 0$, $p_2 > 0, \ldots, p_n > 0$ and $x^2+y^2+\cdots+z^2 < 1$,” *Quart. J. Math.* 2 (1858), 269–300; ibid. 3 (1860), 54–68, 97–108.
☞ 168, 177

[53] L. Schläfli, “Theorie der vielfachen Kontinuität,” in *Gesammelte mathematishe Abhandlungen,* Birkhäuser, Basel, 1950.
☞ 167

[54] J. J. Seidel, “On the volume of a hyperbolic simplex,” *Studia Sci. Math. Hungar.* 21 (1986), no. 1-2, 243–249.
MR 0898862 Zbl 0561.52010 ☞ 171

[55] G. Sforza, “Ricerche di estensionimetria negli spazi metrico-proiettivi,” Memorie Reale Accademia di Scienze, lettere ed arti di Modena, III, VIII (Appendice) (1907), 21–66.
☞ 175

[56] P. Stäckel, “Vizsgálatok az abszolút geometria köréből Bolyai János hátrahagyott irataiban” (”Investigations in absolute geometry in the manuscripts of János Bolyai”)’ Mathematikai és Természettudományi Értesítő XX, Budapest, 1902, 175–186.
☞ 168

[57] P. Stäckel, *Geometrische Untersuchungen von Wolfgang Bolyai und Johann Bolyai* I.–II., B. G. Teubner, Leipzig, 1913.
☞ 168

[58] I. Todhunter, *Spherical trigonometry,* Macmillan and Co., London, 1886.
☞ 152

[59] A. Ushijima, “Volume formula for generalised hyperbolic tetrahedra,” in *Non-Euclidean geometries* (A. Prékopa and E. Molnár, eds.), János Bolyai memorial volume, papers from the International Conference on Hyperbolic Geometry held in Budapest, July 6–12, 2002, Mathematics and Its Applications (New York), 581. Springer, New York, 2006, 249–265.
MR 2191251 Zbl 1096.52006 ☞ 171, 174, 176

[60] J. E. Valentine, “An analogue of Ptolemy’s theorem and its converse in hyperbolic geometry,” *Pacific J. Math.* 34 (1970), 817–825.
MR 0270261 Zbl 0198.25601 ☞ 153

[61] V. V. Varfolomeev, “Inscribed polygons and Heron polynomials,” *Mat. Sb.* 194 (2003), no. 3, 3–24, in Russian; English translation in *Sb. Math.* 194 (2003), no. 3-4, 311–331.
MR 1992154 Zbl 1067.51013 ☞ 154

[62] R. Walter, “Polygons in hyperbolic geometry 1: Rigidity and inversion of the n-inequality,” preprint, 2010.
arXiv:1008.3404v1 [math.MG]
☞ 151, 153

[63] R. Walter, “Polygons in hyperbolic geometry 2: Maximality of area,” preprint, 2010.
arXiv:1008.3821v1 [math.MG]
☞ 151, 152

[64] T. Weszely, *Bolyai János matematikai munkássága* (*Mathematical works of János Bolyai*), Kriterion Könyvkiado, Bukarest, 1981.
☞ 168

[65] L. Wimmer, “Cyclic polygons in non-Euclidean geometry,” *Elem. Math.* 66 (2011), no. 2, 74–82.
MR 2796129 Zbl 1228.51023 ☞ 152, 153

Chapter 12

Statics and kinematics of frameworks in Euclidean and non-Euclidean geometry

Ivan Izmestiev

Contents

1. Introduction 191
2. Kinematics of frameworks 194
3. Statics of frameworks 204
4. Projective statics and kinematics 209
5. Maxwell–Cremona correspodence 216

References 228

1 Introduction

A bar-and-joint framework (framework, for brevity) is made of rigid bars connected at their ends by universal joints. A framework can be constrained to a plane or allowed to move in space. Rigidity of frameworks is a question of practical importance, and its mathematical study goes back to the 19th century. Plate-and-hinge structures such as polyhedra can be represented by bar-and-joint frameworks through replacement of the hinges by bars and rigidifying the plates with the help of diagonals. Thus, rigidity questions for polyhedra belong to the same domain.

There are two ways to approach the rigidity of a framework: through statics, i.e. the ability to respond to exterior loads, and through kinematics, i.e. the absence of deformations. A framework is called *statically rigid* if every system of forces with zero sum and zero moment can be compensated by stresses in the bars of the framework. A framework is called *rigid* if it cannot be deformed while keeping the lengths of all bars, and *infinitesimally rigid* if it cannot be deformed so that the lengths of bars stay constant in the first order. As it turns out, static rigidity is equivalent to infinitesimal rigidity.

The study of statics has a long history. Systems of forces appear in the textbooks of Poinsot [42] and Möbius [37], and the concept of a line-bound force was one of the motivations for Grassmann's introduction of the exterior algebra of a vector space.

Infinitesimal isometric deformations seem to have appeared first in the context of smooth surfaces, see [12] and references therein. In the first half of the 20th century the interest in the isometric deformations was stimulated by the Weyl problem, which asked whether every positively curved Riemannian metric on the 2-sphere can be realized as a smooth convex surface in $\mathbb{R}^3$. This was answered in the affirmative in the 1950s by Nirenberg and by Alexandrov and Pogorelov. The Weyl problem motivated Alexandrov's works on polyhedra, in particular his enhanced version of the Legendre–Cauchy–Dehn rigidity theorem for convex polyhedra. For a survey on rigidity of smooth surfaces see [44], [21], [22], and [20]; for rigidity of frameworks and polyhedra see [9].

The goal of this chapter is to present the fundamental notions and results from the rigidity theory of frameworks in Euclidean space and to extend them to hyperbolic and spherical geometry. Below we state four main theorems whose proofs are given in the subsequent sections.

Theorem A. *A framework in a Euclidean, spherical, or hyperbolic space has equal numbers of kinematic and static degrees of freedom. In particular, infinitesimal rigidity is equivalent to static rigidity.*

The number of static degrees of freedom is the dimension of the vector space of unresolvable loads, that is, of those systems of forces applied to the nodes of the framework which cannot be compensated by stresses in the bars. (A stress in a bar pushes its endpoints apart or pulls them towards each other.) The number of kinematic degrees of freedom is the dimension of the vector space of non-trivial infnitesimal isometric deformations, that is, of deformations which are not induced by an isometry of the ambient space. See Sections 2 and 3 for more detailed definitions and for a proof of Theorem A.

Theorem B (Darboux–Sauer correspondence). *The number of degrees of freedom of a Euclidean framework is a projective invariant. In particular, a framework is infinitesimally rigid if and only if any of its projective images is infinitesimally rigid.*

The projective invariance of static rigidity follows from the interpretation of a line-bound vector (a force) in a d-dimensional Euclidean space as a bivector in $\mathbb{R}^{d+1}$. Linear transformations of $\mathbb{R}^{d+1}$ preserve static dependences; at the same time they generate projective transformations of $\mathbb{RP}^d$. See Section 4.1.

Theorem C (infinitesimal Pogorelov maps). *A hyperbolic or a spherical framework has the same number of kinematic degrees of freedom as its geodesic Euclidean image. In particular, it is infinitesimally rigid if and only if its geodesic Euclidean image is.*

By a geodesic Euclidean image of a hyperbolic framework we mean its representation in a Beltrami–Cayley–Klein model. A geodesic Euclidean image of a spherical framework is its projection from the center of the sphere to an affine hyperplane. Every geodesic map of an open region in the hyperbolic or spherical space into Euclidean space differs from those given above by post-composition with a projective map.

Theorem C is related to Theorem B and is also proved in Section 4.1. In the same section we describe the infinitesimal Pogorelov maps that send the static or kinematic vector spaces of a framework to the corresponding vector spaces of its geodesic image.

While the previous three theorems hold for frameworks of any combinatorics and in spaces of any dimension, the following one is specific to frameworks in dimension 2 whose underlying graphs are planar.

Theorem D (Maxwell–Cremona correspondence). *For a framework on the sphere or in the Euclidean or hyperbolic plane based on a planar graph, the existence of any of the following objects implies the existence of the other two*:

1) *a self-stress,*
2) *a reciprocal diagram,*
3) *a polyhedral lift.*

A self-stress of a framework is a collection of stresses in its bars that resolves a zero load. A reciprocal diagram is essentially a framework whose combinatorics is dual to the combinatorics of a given framework (nodes correspond to faces, bars to bars, and the corresponding bars are perpendicular). Finally, a polyhedral lift in the Euclidean case is a vertical lift to $\mathbb{R}^3$ such that the images of the vertices of every face are coplanar. Definitions of reciprocal diagrams and polyhedral lifts slightly differ in different geometries. Also, Theorem D has various versions all of which are presented in Section 5.

The theory of isometric deformations extends to the smooth case in a straightforward way (and, as we already mentioned, probably preceded the kinematics of frameworks). Accordingly, there are analogs of Theorems B and C for smooth submanifolds of Euclidean, hyperbolic or spherical space. In fact, Theorem B was proved by Darboux for smooth surfaces and only later by Sauer for frameworks [46]. Also Theorem C was first proved by Pogorelov in [41, Chapter 5] for smooth surfaces. On the other hand, a theory of statics for smooth surfaces containing an analog of Theorem A is not fully developed or at least not widely known. (See however the dissertation of Lecornu [31].)

Let us set up the notation used throughout the chapter. In the following, $\mathbb{X}^d$ stands for either $\mathbb{E}^d$ (Euclidean space) or $\mathbb{S}^d$ (spherical space) or $\mathbb{H}^d$ (hyperbolic space). We often view these spaces as subsets of the real vector space $\mathbb{R}^{d+1}$:

$$\begin{aligned}
\mathbb{E}^d &= \{x \in \mathbb{R}^{d+1} \mid x_0 = 1\},\\
\mathbb{S}^d &= \{x \in \mathbb{R}^{d+1} \mid \langle x, x\rangle = 1\},\\
\mathbb{H}^d &= \{x \in \mathbb{R}^{d+1} \mid \langle x, x\rangle_{d,1} = -1, x_0 > 0\}.
\end{aligned}$$

Here $\langle\cdot,\cdot\rangle$ is the Euclidean scalar product, and $\langle\cdot,\cdot\rangle_{d,1}$ is the Minkowski scalar product:

$$\langle x, y\rangle_{d,1} = -x_0 y_0 + x_1 y_1 + \cdots + x_d y_d.$$

Since all formulas involving the scalar products are identical in the spherical and in the hyperbolic case, we omit the subscripts and denote by $\langle\cdot,\cdot\rangle$ both the Euclidean and the Minkowski scalar products. Sometimes we also use $\sin_{\mathbb{X}}$ and $\cos_{\mathbb{X}}$ to denote sin and cos in the spherical and sinh and cosh in the hyperbolic case.

2 Kinematics of frameworks

2.1 Motions

Let Γ be a graph; we denote its vertex set by Γ_0 and its edge set by Γ_1. For the vertices of Γ we use the symbols i, j etc. The edges are unordered pairs of elements of Γ_0, and for brevity we usually write ij instead of $\{i, j\} \in \Gamma_1$.

Definition 2.1. A *framework* in $\mathbb{X}^d$ is a graph Γ together with a map

$$p\colon \Gamma_0 \longrightarrow \mathbb{X}^d, \quad i \longmapsto p_i$$

such that $p_i \neq p_j$ whenever $\{i, j\} \in \Gamma_1$. If $\mathbb{X}^d = \mathbb{S}^d$, then we additionally require $p_i \neq -p_j$ for all $\{i, j\} \in \Gamma_1$.

This is a mathematical abstraction of a bar-and-joint framework, see the introduction. Note that we allow intersections between the edges.

In a framework (Γ, p), every edge receives a non-zero length $\operatorname{dist}(p_i, p_j)$. Two frameworks (Γ, p) and (Γ, p') with the same graph are called *isometric* if they have the same edge lengths: $\operatorname{dist}(p_i, p_j) = \operatorname{dist}(p'_i, p'_j)$ for all $\{i, j\} \in \Gamma_1$. Frameworks with the same graph are called *congruent* if there is an ambient isometry $\Phi \in \operatorname{Isom}(\mathbb{X}^d)$ such that $p'_i = \Phi(p_i)$ for all $i \in \Gamma_0$.

Definition 2.2. A framework (Γ, p) is called *globally rigid* if every framework isometric to (Γ, p) is also congruent to it.

An *isometric deformation* of a framework (Γ, p) is a continuous family of isometric frameworks $(\Gamma, p(t))$, where $t \in (-\epsilon, \epsilon)$ and $p(0) = p$. An isometric deformation is called *trivial* if it is generated by a family of ambient isometries: $p_i(t) = \Phi_t(p_i)$.

Definition 2.3. A framework (Γ, p) is called *rigid* (or locally rigid) if it has no non-trivial isometric deformations. A non-rigid framework is also called *flexible*.

Clearly, global rigidity implies rigidity, but not vice versa. See Figure 1.

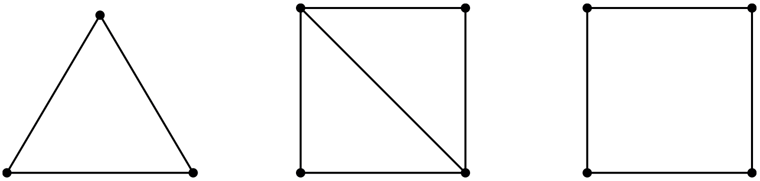


Figure 1. Frameworks in the plane. Left: globally rigid. Middle: rigid but not globally rigid. Right: flexible.

2.2 Infinitesimal motions

Definition 2.4. A *vector field* on a framework (Γ, p) is a map

$$q \colon \Gamma_0 \longrightarrow T\mathbb{X}^d, \quad i \longmapsto q_i$$

such that $q_i \in T_{p_i}\mathbb{X}^d$ for all i. A vector field is called an *infinitesimal isometric deformation* of (Γ, p) if for some (and hence for every) smooth family of frameworks

$(\Gamma, p(t))$ such that

$$p(0) = p, \quad \left.\frac{d}{dt}\right|_{t=0} p_i(t) = q_i \text{ for all } i \in \Gamma_0,$$

we have

$$\left.\frac{d}{dt}\right|_{t=0} \operatorname{dist}(p_i(t), p_j(t)) = 0$$

for all $\{i, j\} \in \Gamma_1$.

Clearly, the infinitesimal isometry condition is equivalent to

$$\langle q_i, e_{ij} \rangle + \langle q_j, e_{ji} \rangle = 0, \tag{1}$$

where $e_{ij} \in T_{p_i}\mathbb{X}^d$ is the unit vector tangent to the geodesic arc $p_i p_j$.

Lemma 2.5. *A vector field q is an infinitesimal isometric deformation of a framework (Γ, p) if and only if*

$$\langle p_i - p_j, q_i - q_j \rangle = 0 \quad \textit{in } \mathbb{E}^d,$$

$$\langle p_i, q_j \rangle + \langle q_i, p_j \rangle = 0 \quad \textit{in } \mathbb{S}^d \textit{ or } \mathbb{H}^d.$$

Here $\langle p_i, q_j \rangle$ means the Euclidean, respectively Minkowski scalar product in $\mathbb{R}^{d+1}$, which makes sense if we identify $T_{p_i}\mathbb{X}^d$ with a linear subspace of $\mathbb{R}^{d+1}$.

Proof. This follows from (1) and

$$e_{ij} = \begin{cases} \dfrac{p_j - p_i}{\|p_j - p_i\|} & \text{in } \mathbb{E}^d, \\ \dfrac{p_j - p_i \cos_{\mathbb{X}} \ell_{ij}}{\sin_{\mathbb{X}} \ell_{ij}} & \text{in } \mathbb{S}^d \text{ and } \mathbb{H}^d, \end{cases} \qquad \square$$

where $= \ell_{ij} = \operatorname{dist}(p_i, p_j)$

An infinitesimal isometric deformation is called *trivial* if there is a Killing field K on $\mathbb{X}^d$ such that $q_i = K(p_i)$ for all i.

Definition 2.6. A framework (Γ, p) is called *infinitesimally rigid*, if it has no non-trivial infinitesimal isometric deformations.

Theorem 2.7. *An infinitesimally rigid framework is rigid.*

For a proof, see [19], [2], and [8].

The converse of Theorem 2.7 is false, see Figure 2.

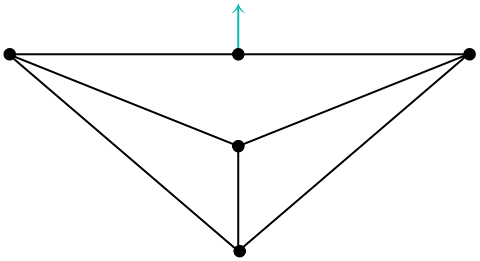


Figure 2. A rigid but infinitesimally flexible framework

Similarly to the example in Figure 2, one can construct a non-trivial infinitesimal isometric deformation for every framework contained in a geodesic subspace of $\mathbb{X}^d$ (provided that the framework has at least $d + 1$ vertices). This is one of the reasons why it is convenient to consider only *spanning* frameworks: those whose vertices are not contained in a geodesic subspace.

Denote the set of all infinitesimal isometric deformations of a framework (Γ, p) by $V(\Gamma, p)$. Due to Lemma 2.5, $V(\Gamma, p)$ is a vector space. The set of trivial infinitesimal isometric deformations is also a vector space; we denote it by $V_0(\Gamma, p)$. If (Γ, p) is spanning, then $\dim V_0(\Gamma, p) = \frac{d(d+1)}{2}$.

Definition 2.8. The dimension of the quotient space $V(\Gamma, p)/V_0(\Gamma, p)$ is called the number of *kinematic degrees of freedom* of a framework (Γ, p).

In particular, infinitesimally rigid frameworks are those with zero kinematic degrees of freedom.

Remark 2.9. Determining whether a framework is flexible is more difficult than determining whether it is infinitesimally flexible: the latter is a linear problem, the former is an algebraic one. Examples of Bricard octahedra and Kokotsakis polyhedra in Section 2.7 illustrate this.

2.3 Point-line frameworks

A *point-line framework* in $\mathbb{R}^2$ associates to every vertex i of Γ either a point p_i or a line l_i in $\mathbb{R}^2$. The edges of Γ correspond to the constraints of the form

$$\operatorname{dist}(p_i, p_j) = \operatorname{dist}(p_i', p_j'),\ \operatorname{dist}(p_i, l_j) = \operatorname{dist}(p_i', l_j'),\ \angle(l_i, l_j) = \angle(l_i', l_j'). \quad (2)$$

For recent works on point-line frameworks see [27] and [15].

In spherical geometry, a point-line framework is equivalent to a standard framework. If we replace every great circle by one of its poles, then the last two constraints in (2) take the form of the first one.

In hyperbolic geometry, the pole of a line is a point in the de Sitter plane (the complement of the disk in the projective model of $\mathbb{H}^2$). Therefore the study of point-line frameworks in $\mathbb{H}^2$ can be reduced to the study of standard frameworks in the hyperbolic-de Sitter plane. Moreover, we can allow ideal points, which means assigning horocycles to some of the vertices of Γ and fixing the point-horocycle, line-horocycle and horocycle-horocycle distances.

2.4 Constraints counting

One can estimate the dimension of the space of non-congruent realizations of a framework by counting the constraints. If $|\Gamma_0| = n$ and $|\Gamma_1| = m$, then there are m equations on dn vertex coordinates. Besides, one has to subtract the dimension of the space of trivial motions, which is $\frac{d(d+1)}{2}$ for spanning frameworks. Thus, generically a framework in $\mathbb{X}^d$ with n vertices and m edges has $dn - m - \frac{d(d+1)}{2}$ degrees of freedom.

Of course, the above arithmetics does not make much sense without the combinatorics (we can put a lot of edges on a subset of the vertices, allowing the other vertices to fly away). Laman [30] has shown that in dimension 2 the arithmetics and combinatorics suffice to characterize generic rigidity. A graph Γ is called a *Laman graph* if $|\Gamma_1| = 2|\Gamma_0| - 3$ and every induced subgraph of Γ with k vertices has at most $2k - 3$ edges.

Theorem 2.10. *A Laman graph is generically rigid, that is, the framework* (Γ, p) *is rigid for almost all* p.

No analog of the Laman condition is known for frameworks in higher dimensions. See [10] for more details on generic rigidity.

If all faces of a 3-dimensional polyhedron homeomorphic to a ball are triangles, then its graph satisfies $|\Gamma_1| = 3|\Gamma_0| - 6$, that is, the above count gives 0 as an estimate for degrees of freedom. Rigidity of polyhedra is discussed in the next section.

2.5 Frameworks and polyhedra

One may try to generalize bar-and-joint frameworks by introducing panel-and-hinge structures: rigid polygons sharing pairs of sides and allowed to freely rotate around these sides, or even more generally n-dimensional "panels" rotating around $(n-1)$-dimensional "hinges." A mathematical model for such an object is called a polyhedron or a polyhedral complex. However, there is a way to replace a polyhedral complex by a framework without changing its isometric deformations (global as well as local and infinitesimal). For this, replace every panel by a complete graph on its vertex set. This "rigidifies" the panels and leaves them the freedom to rotate around the hinges.

A particular class of polyhedral complexes is the one of convex polyhedra. According to the Legendre–Cauchy theorem (see [32] and [7]), a convex polyhedron is globally rigid among convex polyhedra. There are simple examples of convex polyhedra isometric to non-convex ones. By the Dehn theorem [14] (that can also be proved by the Legendre–Cauchy argument), convex 3-dimensional polyhedra are infinitesimally rigid.

The Legendre–Cauchy argument applies to spherical and hyperbolic convex polyhedra as well. This allows to prove the rigidity of convex polyhedra in $\mathbb{X}^d$ for $d > 3$ by induction: the link of a vertex of a d-dimensional convex polyhedron is a $(d-1)$-dimensional spherical polyhedron, and the rigidity of links implies the rigidity of the polyhedron.

A simplicial polyhedron (that is, one all of whose faces are simplices) has the same kinematic properties as its 1-skeleton. In a convex non-simplicial polyhedron we can replace every face by a complete graph as described in the first paragraph; but in fact a much "lighter" framework is enough to keep the polyhedron rigid. It suffices to triangulate every 2-dimensional face in an arbitrary way (without adding new vertices in the interior of the face, vertices on the edges are all right). Again, the Legendre–Cauchy argument ensures the rigidity of all 3-dimensional faces, and the induction applies as in the previous paragraph, see [1, Chapter 10] and [52].

As already indicated, the cone over a framework in $\mathbb{S}^d$ can be viewed as a panel structure (or a framework) in $\mathbb{E}^{d+1}$. Similarly, a framework in $\mathbb{H}^d$ leads to a framework in the $(d+1)$-dimensional Minkowski space.

2.6 Averaging and deaveraging

There is an elegant relation between the infinitesimal and global flexibility. (For smooth surfaces, this idea goes back to the 19^{th}th century.)

Theorem 2.11. 1) DEAVERAGING. *Let* (Γ, p) *be a framework in* $\mathbb{X}^d$ *with a non-trivial infinitesimal isometric deformation* q. *Define two new frameworks* (Γ, p^+) *and* (Γ, p^-) *as follows.*

$$p_i^+ = p_i + q_i, \qquad p_i^- = p_i - q_i, \qquad \textit{for } \mathbb{X} = \mathbb{E},$$

$$p_i^+ = \frac{p_i + q_i}{\|p_i + q_i\|}, \quad p_i^- = \frac{p_i - q_i}{\|p_i - q_i\|}, \quad \textit{for } \mathbb{X} = \mathbb{S} \textit{ or } \mathbb{H},$$

where $\|v\| = \sqrt{|\langle v, v\rangle|}$, *and in the hyperbolic case* $p_i \pm q_i$ *are assumed to be light-like.*

Then the frameworks (Γ, p^+) *and* (Γ, p^-) *are isometric, but not congruent.*

2) AVERAGING. *Let* (Γ, p') *and* (Γ, p'') *be two isometric non-congruent frameworks in* $\mathbb{X}^d$. *Put*

$$p_i = \frac{p_i' + p_i''}{2}, \qquad q_i = \frac{p_i' - p_i''}{2}, \qquad \textit{for } \mathbb{X} = \mathbb{E},$$

$$p_i = \frac{p_i' + p_i''}{\|p_i' + p_i''\|}, \quad q_i = \frac{p_i' - p_i''}{\|p_i' + p_i''\|}, \quad \textit{for } \mathbb{X} = \mathbb{S} \textit{ or } \mathbb{H}.$$

Then q *is a non-trivial infinitesimal isometric deformation of* (Γ, p).

In the deaveraging procedure it might happen that $p_i^+ = p_j^+$ for some $\{i, j\} \in \Gamma_1$, so that p^+ is not a framework. To avoid this, one can replace q by cq for a generic $c \in \mathbb{R}$. In the hyperbolic case, choosing c sufficiently small will make $p_i \pm cq_i$ light-like.

Proof. Formulas of averaging are inverses to those of deaveraging, and both statements can be proved by a direct calculation. Note that in the spherical and the hyperbolic cases we have $\|p_i + q_i\| = \|p_i - q_i\|$ due to $\langle p_i, q_i\rangle = 0$. Also q is non-trivial if and only if it changes the distance in the first order between some p_i and p_j not connected by an edge. One can check that this is equivalent to $\operatorname{dist}(p_i^+, p_j^+) \neq \operatorname{dist}(p_i^-, p_j^-)$. □

2.7 Examples

In Section 2.4 we spoke about generically rigid graphs. The most interesting examples of flexible frameworks are special realizations of generically rigid graphs.

Example 2.12 (a planar framework with 9 edges on 6 vertices). The frameworks on Figure 3 (which are combinatorially equivalent) are infinitesimally flexible if and only if the lines a, b, c are concurrent, that is, meeting at a point or parallel. This can be proved with the help of the Maxwell–Cremona correspondence, see Example 5.4.

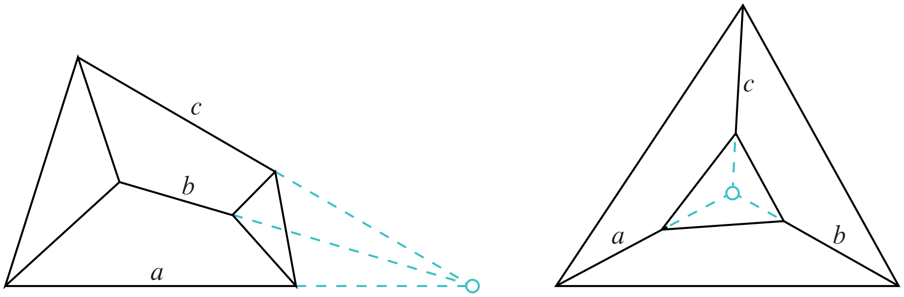


Figure 3. Infinitesimally flexible frameworks in the plane

Example 2.13 (another planar framework with 9 edges on 6 vertices). A framework based on the bipartite graph $K_{3,3}$ is infinitesimally flexible if and only if its vertices lie on a (possibly degenerate) conic, see Figure 4. For a proof see [5] and [51]. On the left-hand side, the vertices lie on a circle; the arrows indicate a non-trivial infinitesimal isometric deformation.

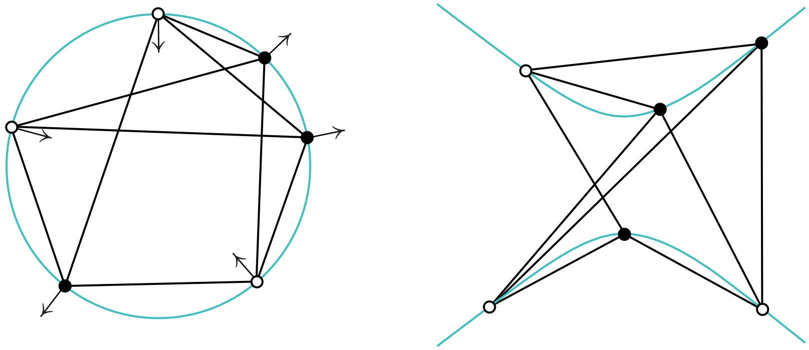


Figure 4. Infinitesimally flexible frameworks in the plane

The conditions in the above two examples are projectively invariant. Besides, the same criteria hold for frameworks on the sphere or in the projective plane. (Three lines in $\mathbb{H}^2$ are called concurrent if they meet at a hyperbolic, ideal, or de Sitter point.) A non-Euclidean conic is one that is depicted by an affine conic in a geodesic model of $\mathbb{S}^2$ or $\mathbb{H}^2$, see [23].

Example 2.14 (Bricard's octahedra and Gaifullin's cross-polytopes). Flexible octahedra (with intersecting faces) were discovered and classified by Bricard [6], see also [49] and [38]. A higher-dimensional analog of the octahedron is called cross-polytope. Recently, flexible cross-polytopes in $\mathbb{X}^d$ were classified by Gaifullin [18].

Example 2.15 (infinitesimally flexible octahedra). While the description and classification of flexible octahedra requires quite some work, infinitesimally flexible octahedra can be described in a simple and elegant way.

> Color the faces of an octahedron white and black so that adjacent faces receive different colors. An octahedron is infinitesimally flexible if and only if the planes of its four white faces meet at a point (which may lie at infinity). As a consequence, the planes of the white faces meet if and only if those of the black faces do.

This theorem was proved independently by Blaschke and Liebmann [4] and [33]. The configuration is related to the so-called Möbius tetrahedra: a pair of mutually inscribed tetrahedra, [36].

Figure 5 shows two examples of infinitesimally flexible octahedra. The one on the left is a special case of the Schoenhardt octahedron [47]; its bases are regular triangles, and the orthogonal projection of one base to the other makes the triangles concentric with pairwise perpendicular edges.

Theorem B implies that infinitesimally flexible octahedra in $\mathbb{S}^3$ and $\mathbb{H}^3$ are characterized by the same criterion as those in $\mathbb{E}^3$. In hyperbolic space, the intersection point of four planes may be ideal or hyperideal. In fact, even the vertices of an octahedron may be ideal or hyperideal. Infinitesimally flexible hyperbolic octahedra were used in [25] to construct simple examples of infinitesimally flexible hyperbolic cone-manifolds.

Example 2.16 (Jessen's icosahedron and its relatives). In the xy-plane of $\mathbb{R}^3$, take the rectangle with vertices $(\pm 1, \pm, t, 0)$, where $0 < t < 1$. Take two other rectangles, obtained from this one by 120° and 240° rotations around the $x = y = z$ line (which results in cyclic permutations of the coordinates). The convex hull of the twelve vertices of these rectangles is an icosahedron (a regular one for $t = \frac{\sqrt{5}-1}{2}$). Among the edges of this icosahedron are the short sides of the rectangles.

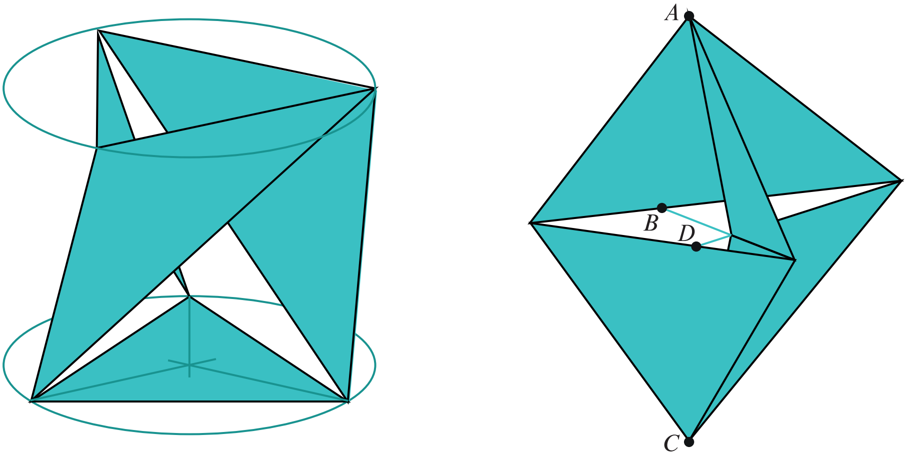


Figure 5. Infinitesimally flexible octahedra. On the right, the points A, B, C, D must be coplanar.

Modify the 1-skeleton by removing the short sides of rectangles (like the one joining $(1, t, 0)$ with $(1, -t, 0)$) and inserting the long sides (like the one joining $(1, t, 0)$ with $(-1, t, 0)$). The resulting framework $p(t)$ is the 1-skeleton of a non-convex icosahedron. Jessen [28] gave the $t = \frac{1}{2}$ non-convex icosahedron as an example of a closed polyhedron with orthogonal pairs of adjacent faces, but different from the cube. See Figure 6.

The framework $p(t)$ has two sorts of edges: the long sides of the rectangles, which have length 2, and the sides of eight equilateral triangles, which have length $\sqrt{2(t^2 - t + 1)}$. It follows that the frameworks $p(t)$ and $p(1-t)$ are isometric. Note that $p(0)$ collapses to an octahedron: the map $p(t)\colon \Gamma_0 \to \mathbb{R}^3$ sends the vertices of the icosahedral graph to the vertices of a regular octahedron by identifying them in pairs; there are three pairs of edges that are mapped to three diagonals of the octahedron. At the same time, $p(1)$ is the graph of the cuboctahedron with square faces subdivided in a certain way.

Since the average of $p(t)$ and $p(1 - t)$ (in the sense of Section 2.6) is $p\left(\frac{1}{2}\right)$, it follows that Jessen's icosahedron is infinitesimally flexible.

Theorem B implies that there are spherical and hyperbolic analogs of this construction.

Example 2.17 (Kokotsakis polyhedra). A Kokotsakis polyhedron with an n-gonal base is a panel structure made of a rigid n-gon, and n quadrilaterals attached to its edges, and n triangles attached between the quadrilaterals, see Figure 7, left. Generically, a Kokotsakis polyhedron is rigid; it is flexible for certain symmetric configurations, see Figure 7, right.

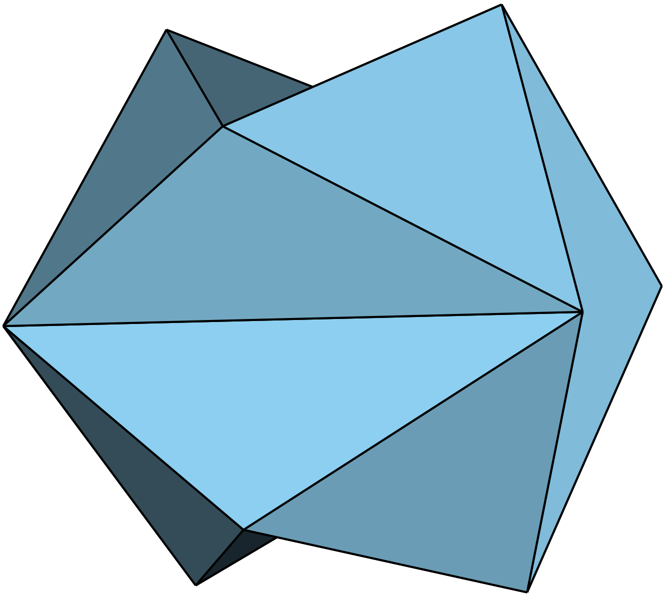


Figure 6. Jessen's orthogonal and infinitesimally flexible icosahedron

Especially interesting are the polyhedra with a quadrangular base, because of their relation to quad-surfaces (polyhedral surfaces made of quadrilaterals with four quadrilaterals around each vertex). A quad-surface is (infinitesimally) flexible if and only if all Kokotsakis polyhedra around its faces are. A famous example of a flexible quad-surface is the Miura-ori [39].

A characterization of infinitesimally flexible Kokotsakis polyhedra was given by Kokotsakis in [29]; several flexible examples were constructed in [45] and [29]. A complete classification of flexible polyhedra with a quadrangular base is given in [26].

3 Statics of frameworks

3.1 Euclidean statics

In the statics of a rigid body, a force is represented as a line-bound vector: moving the force vector along the line it spans does not change its action on a rigid body.

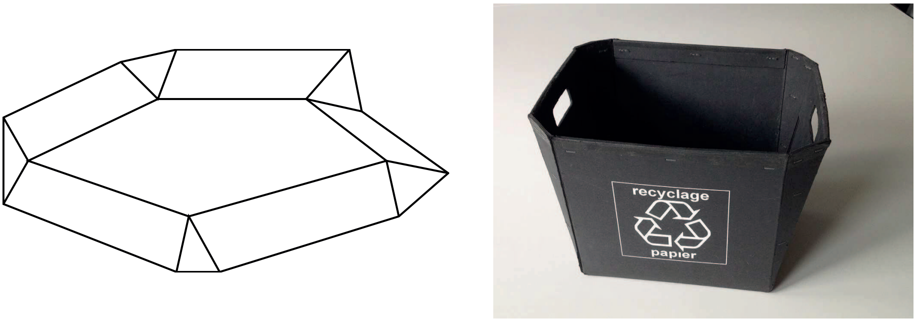


Figure 7. Kokotsakis polyhedra

Definition 3.1. A *force* in a Euclidean space is a pair (p, f) with $p \in \mathbb{E}^d$, $f \in \mathbb{R}^d$. A *system of forces* is a formal sum $\sum_i (p_i, f_i)$ that may be transformed according to the following rules.

0) A force with a zero vector is a zero force:

$$(p, 0) \sim 0.$$

1) Forces at the same point can be added and scaled as usual:

$$\lambda_1(p, f_1) + \lambda_2(p, f_2) \sim (p, \lambda_1 f_1 + \lambda_2 f_2).$$

2) A force may be moved along its line of action:

$$(p, f) \sim (p + \lambda f, f).$$

One may check from this definition that systems of forces form a vector space of dimension $\frac{d(d+1)}{2}$.

In $\mathbb{E}^2$, any system of forces is equivalent either to a single force or to a so called "force couple" $(p_1, f) + (p_2, -f)$, where the vector f is not parallel to the line through p_1 and p_2.

Definition 3.2. A *load* on a Euclidean framework (Γ, p) is a map

$$\begin{aligned} f\colon \Gamma_0 &\longrightarrow \mathbb{R}^d, \\ i &\longmapsto f_i. \end{aligned}$$

A load is called an *equilibrium load* if the system of forces $\sum_{i \in \Gamma_0} (p_i, f_i)$ is equivalent to a zero force.

A rigid body responds to an equilibrium load by interior stresses that cancel the forces of the load. This motivates the following definition.

Definition 3.3. A *stress* on a framework (Γ, p) is a map

$$\begin{aligned} w\colon \Gamma_1 &\longrightarrow \mathbb{R}, \\ ij &\longmapsto w_{ij} = w_{ji}. \end{aligned}$$

The stress w is said to *resolve* the load f if

$$f_i = \sum_{j \in \Gamma_0} w_{ij}(p_j - p_i) \quad \text{for all } i \in \Gamma_0, \tag{3}$$

where we put $w_{ij} = 0$ for all $ij \notin \Gamma_1$.

We denote the vector space of equilibrium loads by $F(\Gamma, p)$, and the vector space of resolvable loads by $F_0(\Gamma, p)$. It is easy to see that every resolvable load is an equilibrium load: $F_0(\Gamma, p) \subset F(\Gamma, p)$.

Definition 3.4. The dimension of the quotient space $F(\Gamma, p)/F_0(\Gamma, p)$ is called the number of *static degrees of freedom* of the framework (Γ, p).

The framework (Γ, p) is called *statically rigid* if it has zero static degrees of freedom, i.e. if every equilibrium load can be resolved.

3.2 Non-Euclidean statics

Definition 3.5. Let $\mathbb{X}^d = \mathbb{S}^d$ or $\mathbb{H}^d$. A *force* in $\mathbb{X}^d$ is an element of the tangent bundle $T\mathbb{X}^d$. We write it as a pair (p, f) with $p \in \mathbb{X}^d$ and $f \in T_p\mathbb{X}^d$.

A *system of forces* is a formal sum of forces that may be transformed according to the rules of Definition 3.1, where the formula in the rule 2) is replaced by $(p, f) \sim (\exp_p(\lambda f), \tau(f))$ with $\tau(f)$ being the result of the parallel transport of f along the geodesic from p to $\exp_p(\lambda f)$.

A system of forces on $\mathbb{S}^2$ is always equivalent to a single force; a system of forces on $\mathbb{H}^2$ is equivalent to either a single force, or an ideal force couple or a hyperideal force couple.

Definition 3.6. A load on a framework (Γ, p) in $\mathbb{X}^d$ is a map

$$f\colon \Gamma_0 \longrightarrow T\mathbb{X}^d, \quad f_i \in T_{p_i}\mathbb{X}^d.$$

A load is called an *equilibrium load* if the system of forces $\sum_{i \in \Gamma_0}(p_i, f_i)$ is equivalent to a zero force.

In the above definitions, $\mathbb{X}^d$ can also stand for $\mathbb{E}^d$. The canonical isomorphisms $T_x\mathbb{E}^d \cong \mathbb{R}^d$ result in simplified formulations given in the preceding section.

As in the Euclidean case, a stress on a framework in $\mathbb{X}^d$ is a map $w\colon \Gamma_1 \to \mathbb{R}$. A stress w *resolves a load* f if

$$f_i = \sum_{j \in \Gamma_0} w_{ij} \ell_{ij} e_{ij}.$$

The following lemma gives an alternative description of the stress resolution.

Lemma 3.7. *A stress w resolves a load f on a framework (Γ, p) in $\mathbb{X}^d = \mathbb{S}^d$ or $\mathbb{H}^d$ if and only if for every $i \in \Gamma_0$ we have*

$$f_i - \sum_{j \in \Gamma_0} \lambda_{ij} p_j \parallel p_i,$$

where

$$\lambda_{ij} = w_{ij} \frac{\ell_{ij}}{\sin_{\mathbb{X}} \ell_{ij}}.$$

Here $f_i, p_i \in \mathbb{R}^{d+1}$ via $\mathbb{X}^d \subset \mathbb{R}^{d+1}$, and $v \parallel w$ means that the vectors v and w are linearly dependent.

Proof. Follows from the formula for e_{ij} given in the proof of Lemma 2.5.

$$p_j - \cos_{\mathbb{X}} \operatorname{dist}(p_i, p_j) p_i = p_j - \langle p_i, p_j \rangle p_i = \sin_{\mathbb{X}} \operatorname{dist}(p_i, p_j) \cdot e_{ij}. \qquad \square$$

3.3 Equivalence of static and infinitesimal rigidity

Define a pairing between vector fields and loads on a framework (Γ, p):

$$\langle q, f \rangle = \sum_{i \in \Gamma_0} \langle q_i, f_i \rangle. \tag{4}$$

This pairing is non-degenerate and therefore induces a duality between the space of vector fields and the space of loads.

Lemma 3.8 (principles of virtual work). *Under the pairing* (4),

1) *the space of infinitesimal motions is the annihilator of the space of resolvable loads*:

$$V(\Gamma, p) = F_0(\Gamma, p)^\circ;$$

2) *the space of trivial infinitesimal motions is the annihilator of the space of equilibrium loads*:

$$V_0(\Gamma, p) = F(\Gamma, p)^\circ.$$

A proof in the Euclidean case can be found in [24]; it transfers to the spherical and the hypebolic cases.

As a consequence, the pairing (4) induces an isomorphism

$$V(\Gamma, p)/V_0(\Gamma, p) \cong (F(\Gamma, p)/F_0(\Gamma, p))^* \tag{5}$$

which implies Theorem A.

The statics of a Euclidean framework is formulated in purely linear terms: loads and stresses on a framework correspond to loads and stresses on its affine image. Together with Theorem A this leads to the following conclusion, which is a special case of Theorem B.

Corollary 3.9. *The number of kinematic degrees of freedom of a Euclidean framework is an affine invariant. In particular, an affine image of an infinitesimally rigid framework is infinitesimally rigid.*

Definition 3.10. The *rigidity matrix* of a Euclidean framework (Γ, p) is a $\Gamma_1 \times \Gamma_0$ matrix with vector entries:

$$\mathcal{R}(\Gamma, p) = \begin{array}{cc} & \begin{array}{ccc} & i & \end{array} \\ ij & \begin{pmatrix} & \vdots & \\ \cdots & p_i - p_j & \cdots \\ & \vdots & \end{pmatrix} \end{array}.$$

It has the pattern of the edge-vertex incidence matrix of the graph Γ, with $p_i - p_j$ on the intersection of the row ij and the column i.

The rows of $\mathcal{R}(\Gamma, p)$ span the space $F_0(\Gamma, p)$. The following proposition is a reformulation of the first principle of virtual work.

Lemma 3.11. *Consider* $\mathcal{R}(\Gamma, p)$ *as the matrix of a map* $(\mathbb{R}^d)^{\Gamma_0} \to \mathbb{R}^{\Gamma_1}$. *Then the following holds*:

$$\ker \mathcal{R}(\Gamma, p) = V(\Gamma, p);$$

$$\operatorname{im} \mathcal{R}(\Gamma, p)^\top = F_0(\Gamma, p).$$

Corollary 3.12. *A framework* (Γ, p) *is infinitesimally rigid if and only if*

$$\operatorname{rk} \mathcal{R}(\Gamma, p) = d\,|\Gamma_0| - \binom{d+1}{2}.$$

4 Projective statics and kinematics

4.1 Projective statics

For $\mathbb{X}^d = \mathbb{E}^d$, $\mathbb{S}^d$ or $\mathbb{H}^d$, associate to a force (p, f) in $\mathbb{X}^d$ a bivector in $\mathbb{R}^{d+1}$:

$$(p, f) \longmapsto p \wedge f. \tag{6}$$

We use the canonical embeddings $\mathbb{X}^d \subset \mathbb{R}^{d+1}$ that allow to view a point p and a vector f as vectors in $\mathbb{R}^{d+1}$.

Lemma 4.1. *The map* (6) *extends to an isomorphism between the space of systems of forces on* $\mathbb{X}^d$ *and the second exterior power* $\Lambda^2(\mathbb{R}^{d+1})$.

The equivalence relations from Definition 3.1 ensure that a linear extension is well-defined. For a proof of its bijectivity, see [24].

The above observation motivates the following definitions.

Definition 4.2. A *projective framework* is a graph Γ together with a map

$$\pi\colon \Gamma_0 \longrightarrow \mathbb{RP}^d, \quad i \longmapsto \pi_i,$$

such that $\pi_i \neq \pi_j$ for $ij \in \Gamma_1$.

We say that $\phi \in \Lambda^2(\mathbb{R}^{d+1})$ is *divisible* by a vector v if $\phi = v \wedge w$ for some vector w. Similarly, we say that ϕ is divisible by $\pi \in \mathbb{RP}^d$ if ϕ is divisible by a representative of π.

Definition 4.3. A *load on a projective framework* (Γ, π) is a map

$$\phi\colon \Gamma_0 \longrightarrow \Lambda^2(\mathbb{R}^{d+1}), \quad i \longmapsto \phi_i,$$

that sends every vertex i to a bivector divisible by π_i. An *equilibrium load* is one that satisfies

$$\sum_{i \in \Gamma_0} \phi_i = 0.$$

Definition 4.4. Denote by Γ_1^{or} the set of oriented edges of the graph Γ. A *stress on a projective framework* (Γ, π) is a map

$$\Omega\colon \Gamma_1^{\mathrm{or}} \longrightarrow \Lambda^2(\mathbb{R}^{d+1}), \quad (i, j) \longmapsto \omega_{ij}$$

such that ω_{ij} is divisible by both π_i and π_j, and $\omega_{ij} = -\omega_{ji}$.

A stress Ω is said to *resolve a load* ϕ if

$$\phi_i = \sum_{j \in \Gamma_0} \omega_{ij}.$$

The *projectivization* of a framework (Γ, p) in $\mathbb{X}^d$ is obtained by composing p with the inclusion $\mathbb{X}^d \subset \mathbb{R}^{d+1}$ and the projection $\mathbb{R}^{d+1} \setminus \{0\} \to \mathbb{RP}^d$. The following lemma is straightforward.

Lemma 4.5. *The map* (6) *sends bijectively the equilibrium, respectively resolvable, loads on a framework in* $\mathbb{X}^d$ *to the equilibrium, respectively resolvable, loads on its projectivization.*

Theorems B and C are immediate corollaries of Lemma 4.5.

Proof of Theorem B. Two frameworks in $\mathbb{E}^d$ are projective images of one another if and only if their projectivizations are related by a linear isomorphism of $\mathbb{R}^{d+1}$. A linear map sends equilibrium loads to equilibrium ones, and resolvable to resolvable ones. ☐

It seems that Theorem B was first proved by Rankine [43] in 1863. He stated that static rigidity is a projective invariant but did not give the details, just saying that "[…] theorems discovered by Mr. Sylvester […] obviously give at once the solution of the question." The first detailed accounts are [33] (for a special case $|\Gamma_1| = d|\Gamma_0| - \frac{d(d+1)}{2}$) and [46].

Proof of Theorem C. A Euclidean framework and its geodesic spherical or hyperbolic image have the same projectivizations. Hence the maps (6) yield an isomorphism between the spaces of their equilibrium/resolvable loads. ☐

4.2 Static and kinematic Pogorelov maps

Let a framework (Γ, p) in $\mathbb{E}^d$ and a projective map $\Phi\colon \mathbb{RP}^d \to \mathbb{RP}^d$ be given such that the image of $\Phi \circ p$ is contained in $\mathbb{E}^d$. (Here $\mathbb{RP}^d$ is a projective completion of $\mathbb{E}^d$). Lemma 4.5 does not only show that the spaces of equilibrium modulo resolvable loads of (Γ, p) and $(\Gamma, \Phi \circ p)$ have the same dimension, but also establishes a canonical up to a scalar factor isomorphism between these spaces. Through the static-kinematic duality from Section 3.3 this also yields an isomorphism between the spaces of infinitesimally isometric modulo trivial motions.

The situation is similar with the geodesic correspondence between frameworks in different geometries. The kinematic isomorphisms were described by Pogorelov in [41, Chapter 5] together with the maps that associate to a pair of isometric polyhedra in one geometry a pair of isometric polyhedra with the same combinatorics in the other geometry (this is related to the kinematic isomorphism via the averaging procedure, see Section 2.6). We will use the term *Pogorelov maps* in each of the above situations.

Definition 4.6. Let $X \subset \mathbb{X}^d$ and $Y \subset \mathbb{Y}^d$, $\mathbb{X}, \mathbb{Y} \in \{\mathbb{E}, \mathbb{S}, \mathbb{H}\}$, and let $\Phi\colon X \to Y$ be a geodesic map. A fiberwise linear map $\Phi^{\mathrm{stat}}\colon TX \to TY$ with $\Phi^{\mathrm{stat}}(T_pX) \subset T_{\Phi(p)}Y$ is called a *static Pogorelov map* associated with Φ if for every framework (Γ, p) in X the following two conditions are satisfied:

- a load f on (Γ, p) is in equilibrium if and only if the load $\Phi^{\mathrm{stat}} \circ f$ on the framework $(\Gamma, \Phi \circ p)$ is in equilibrium;
- a load f on (Γ, p) is resolvable if and only if the load $\Phi^{\mathrm{stat}} \circ f$ on the framework $(\Gamma, \Phi \circ p)$ is resolvable.

A fiberwise linear map $\Phi^{\mathrm{kin}}\colon TX \to TY$ with $\Phi^{\mathrm{kin}}(T_pX) \subset T_{\Phi(p)}Y$ is called a *kinematic Pogorelov map* associated with Φ if for every framework (Γ, p) in X the following two conditions are satisfied:

- a vector field q on (Γ, p) is an infinitesimal isometric deformation if and only if the vector field $\Phi^{\mathrm{kin}} \circ q$ on $(\Gamma, \Phi \circ p)$ is an infinitesimal isometric deformation;
- a vector field q on (Γ, p) is a trivial infinitesimal isometric deformation if and only if the vector field $\Phi^{\mathrm{kin}} \circ q$ on $(\Gamma, \Phi \circ p)$ is a trivial infinitesimal isometric deformation.

Remark 4.7. The last condition on a kinematic Pogorelov map means that Φ^{kin} sends Killing fields on X to Killing fields on Y. For an intrinsic approach to the Pogorelov maps defined for Riemannian metrics with the same geodesics, see [16, Section 4.3].

Lemma 4.8. *If* Φ^{stat} *is a static Pogorelov map associated with* Φ*, then* $((\Phi^{\mathrm{stat}})^{-1})^*$ *is a kinematic Pogorelov map associated with* Φ*.*

Proof. This follows from

$$\langle((\Phi^{\mathrm{stat}})^{-1})^*(q), \Phi^{\mathrm{stat}}(f)\rangle = \langle q, (\Phi^{\mathrm{stat}})^{-1} \circ \Phi^{\mathrm{stat}}(q)\rangle = \langle q, f\rangle$$

and from Lemma 3.8. ☐

4.3 Pogorelov maps for affine and projective transformations

Theorem 4.9. *Let* $\Phi \colon \mathbb{E}^d \to \mathbb{E}^d$ *be an affine transformation with the linear part* $A = d\Phi \in \mathrm{GL}(n, \mathbb{R})$*. Then*

$$\Phi^{\mathrm{stat}} = A, \quad \Phi^{\mathrm{kin}} = (A^{-1})^*$$

are static and kinematic Pogorelov maps for Φ*.*

Proof. The equivalence relation in Definition 3.1 is affinely invariant. Therefore f is an equilibrium load on (Γ, p) if and only if $A \circ f$ is an equilibrium load on $(\Gamma, \Phi \circ p)$. When (Γ, p) is transformed by Φ, the right-hand side of (3) is transformed by A. Therefore a stress that resolves f also resolves $A \circ f$. ☐

Theorem 4.10. *Let*

$$\Phi \colon \mathbb{E}^d \setminus L \longrightarrow \mathbb{E}^d \setminus L'$$

be a projective transformation, where L *is the hyperplane sent to infinity, and* L' *the image of the hyperplane at infinity. Denote by* $h_L(p)$ *the distance from a point* $p \in \mathbb{E}^d$ *to the hyperplane* L*. Then*

$$\Phi_p^{\mathrm{stat}} = h_L^2(p) \cdot d\Phi_p, \quad \Phi_p^{\mathrm{kin}} = h_L^{-2}(p) \cdot ((d\Phi_p)^*)^{-1}$$

are static and kinematic Pogorelov maps for Φ*.*

Proof. A projective map Φ consists of a linear transformation $M \in \mathrm{GL}(d+1, \mathbb{R})$ restricted to $\mathbb{E}^d$ followed by the central projection from the origin to $\mathbb{E}^d$. We need to compose the map (6) with $M_* \colon \Lambda^2(\mathbb{R}^{d+1}) \to \Lambda^2(\mathbb{R}^{d+1})$ and then with the inverse of (6).

The map (6) followed by M_* transforms a force (p, v) as follows:

$$(p, v) \longmapsto p \wedge v = p \wedge (p + v) \longmapsto M(p) \wedge M(p + v).$$

We have

$$M(p) = \frac{h_{M(L)}(M(p))}{\operatorname{dist}(\mathbb{E}^d \cap M(\mathbb{E}^d), M(L))} \Phi(p) = c \cdot h_L(p) \cdot \Phi(p)$$

for some $c \in \mathbb{R}$, where the distances are taken with a sign, see Figure 8. It follows that

$$M(p) \wedge M(p + v) = c^2 \cdot h_L(p) \cdot h_L(p + v) \cdot \Phi(p) \wedge \Phi(p + v).$$

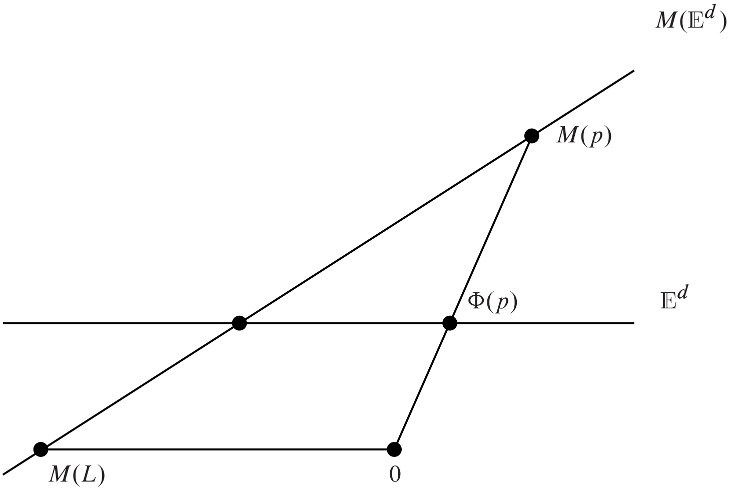


Figure 8. Computing the Pogorelov map for a projective transformation

Applying the inverse of (6) we see that the vector v at p is transformed into the vector

$$c^2 \cdot h_L(p) \cdot h_L(p + v) \cdot (\Phi(p + v) - \Phi(p))$$

at $\Phi(p)$. By construction, this transformation is linear in v. Therefore it does not change if we replace v by tv and take the derivative with respect to t at $t = 0$. This derivative equals $c^2 h_L^2(p) d\Phi_p(v)$. This proves the formula for Φ^{stat}. The formula for Φ^{kin} follows from Lemma 4.8. □

4.4 Pogorelov maps for geodesic projections of $\mathbb{S}^d$ and $\mathbb{H}^d$

Theorem 4.11. *Let* $G\colon \mathbb{E}^d \to X$ *be the projection from the origin of* $\mathbb{R}^{d+1}$*, where* $X = \mathbb{S}^d_+$ *or* $X = \mathbb{H}^d$*.*

Then the Pogorelov maps for a Euclidean framework (Γ, p) *and its spherical, respectively hyperbolic, image* $(\Gamma, G \circ p)$ *are given by*

$$G_p^{\mathrm{stat}} = \|p\| \cdot dG_p, \quad G_p^{\mathrm{kin}} = \frac{1}{\|p\|}(dG_p^*)^{-1}.$$

Here $\|\cdot\| = \sqrt{|\langle\cdot,\cdot\rangle|}$*.*

Note that in the spherical case at the point e_0 (the tangency point of X with $\mathbb{E}^d$) we have $G_{e_0}^{\mathrm{stat}} = dG_{e_0}$.

Proof. To compute the image of $v \in T_p\mathbb{E}^d$ under the differential dG_p, project the geodesic $p + tv$ in $\mathbb{E}^d$ to X. Then $dG_p(v)$ is the velocity vector of the projected curve at $t = 0$, see Figure 9, top, that illustrates the case of the sphere. On the other hand, the image of v under the static Pogorelov map is determined by

$$G(p) \wedge G_p^{\mathrm{stat}}(v) = p \wedge v.$$

Hence both $dG_p(v)$ and $G_p^{\mathrm{stat}}(v)$ are linear combinations of p and v tangent to $\mathbb{S}^d$. It follows that these two vectors are collinear:

$$G_p^{\mathrm{stat}}(v) = \lambda(p,v) \cdot dG_p(v), \quad \lambda(p,v) \in \mathbb{R}.$$

If the images of every vector under two linear maps are collinear, then these maps are scalar multiples of each other. Thus λ depends on p only:

$$G_p^{\mathrm{stat}} = \lambda(p) \cdot dG_p.$$

For small t, the ratio of the areas of the triangles on Figure 9, top, is equal to $\|p\|$. Hence

$$G(p) \wedge dG_p(v) = \frac{1}{\|p\|} p \wedge v,$$

which implies the first formula of the theorem. The second formula follows from the duality between infinitesimal deformations and loads. $\square$

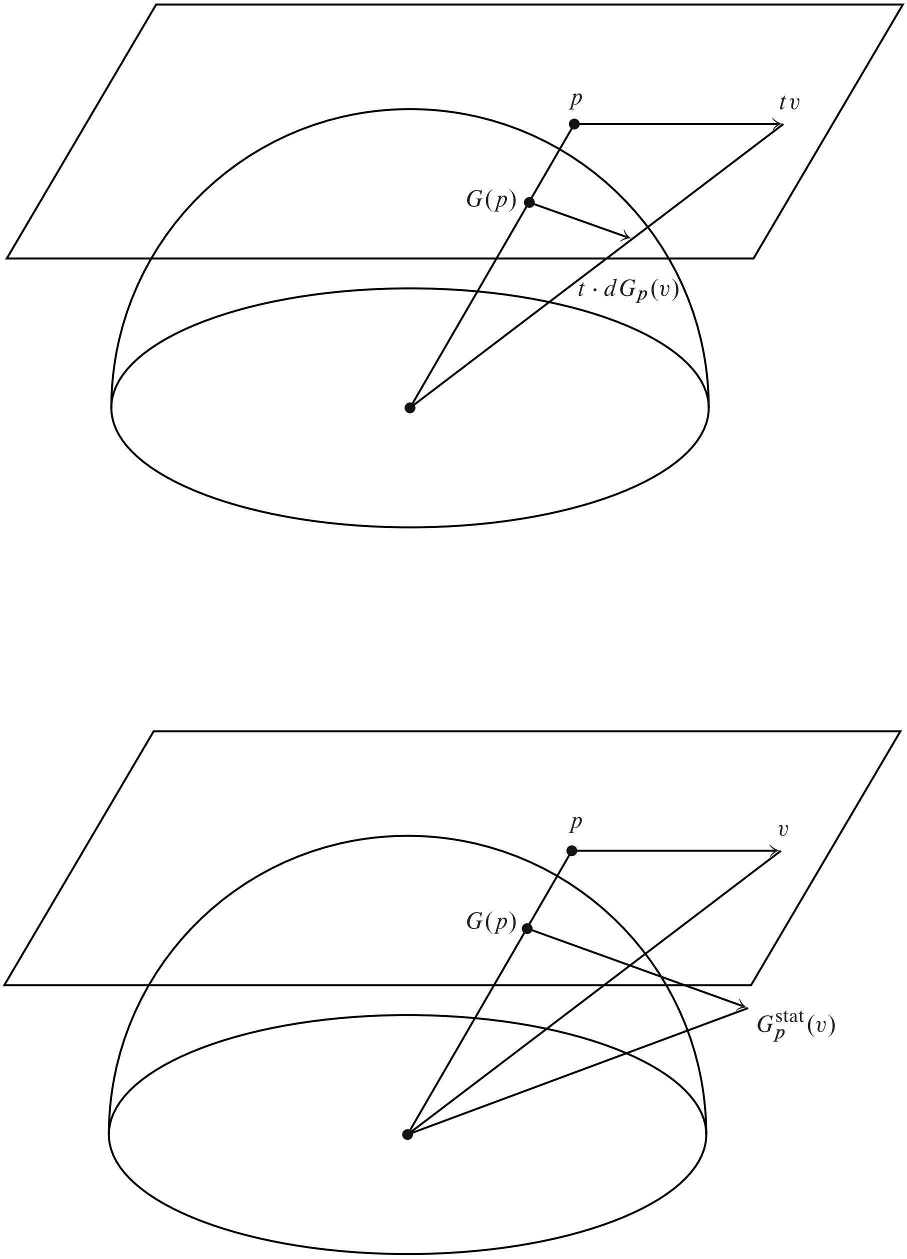


Figure 9. Computing the Pogorelov map for a geodesic projection to the sphere

5 Maxwell–Cremona correspodence

5.1 Planar 3-connected graphs, polyhedra, and duality

A graph is called 3-*connected* if it is connected, has at least 4 vertices, and remains connected after removal of any two of its vertices. In particular, every vertex of a 3-connected graph has degree at least 3.

Planar 3-connected graphs have very nice properties. First, by a result of Whitney [54], their embeddings into $\mathbb{S}^2$ split in two isotopy classes that differ by an orientation-reversing diffeomorphism of $\mathbb{S}^2$. Second, by the Steinitz theorem (see [50] and [55]), a graph is planar and 3-connected if and only if it is isomorphic to the skeleton of some convex 3-dimensional polyhedron. Whitney's theorem implies that for a planar 3-connected graph Γ there is a well-defined set of *faces* Γ_2. Geometrically, a face is a connected component of $\mathbb{S}^2 \setminus \phi(\Gamma)$, where ϕ is an embedding of Γ; combinatorially it is the set of vertices on the boundary of such a component. We call (α, i) with $\alpha \in \Gamma_2$, $i \in \Gamma_0$ and $i \in \alpha$ an *incident pair*. The choice of an isotopy class of an embedding $\Gamma \to \mathbb{S}^2$ and of an orientation of $\mathbb{S}^2$ induces a cyclic order on the set of vertices incident to a face.

The *dual graph* Γ^* of a planar 3-connected graph Γ can be constructed from an embedding $\Gamma \to \mathbb{S}^2$ by choosing a point inside every face and joining every pair of points whose corresponding faces share an edge. The graph Γ^* is also planar and 3-connected, and its dual is again Γ. If an edge ij of Γ separates the faces α and β, then we say that $(\alpha\beta, ij)$ is a *dual pair of edges*. Choose an isotopy class of embeddings $\Gamma \to \mathbb{S}^2$ and fix an orientation of $\mathbb{S}^2$. Then we say that the pair $(\alpha\beta, ij)$ is *consistently oriented* if the face α lies on the right from the edge ij directed from i to j, see Figure 10. Changing the order of i and j or of α and β transforms an inconsistently oriented pair into a consistently oriented one.

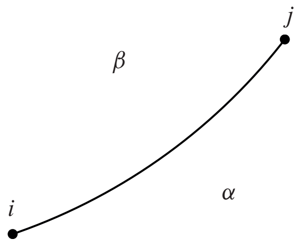


Figure 10. A consistently oriented dual pair

5.2 Maxwell–Cremona theorem

For convenience we identify in this section $\mathbb{E}^2$ with $\mathbb{R}^2$ by choosing an origin.

Definition 5.1. Let (Γ, p) be a framework in $\mathbb{R}^2$ with a planar 3-connected graph Γ. A *reciprocal diagram* for (Γ, p) is a framework (Γ^*, m) such that dual edges are perpendicular to each other:

$$m_\beta - m_\alpha \perp p_j - p_i$$

whenever the edge ij of Γ separates the faces α and β.

Definition 5.2. Let (Γ, p) be a framework in $\mathbb{R}^2$ with a planar 3-connected graph Γ and such that for every face $\alpha \in \Gamma_2$ the points $\{p_i \mid i \in \alpha\}$ are not collinear. A *vertical polyhedral lift* of (Γ, p) is a map $\tilde{p} \colon \Gamma_0 \to \mathbb{R}^3$ such that

1) $\mathrm{pr}_\perp \circ \tilde{p} = p$, where $\mathrm{pr}_\perp \colon \mathbb{R}^3 \to \mathbb{R}^2$ is the orthogonal projection;
2) for every face α of (Γ, p) the points $\{\tilde{p}_i \mid i \in \alpha\}$ are coplanar;
3) the planes of the adjacent faces differ from each other.

A *radial polyhedral lift* of (Γ, p) is a map $\tilde{p} \colon \Gamma_0 \to \mathbb{R}^3$ that satisfies the above conditions with 1) replaced by

1′) $\mathrm{pr}_a \circ \tilde{p} = p$, where $\mathrm{pr}_a \colon \mathbb{R}^3 \setminus \{a\} \to \mathbb{R}^2$ is the radial projection from a point $a \notin \mathbb{R}^2$.

It turns out that reciprocal diagrams are related to polyhedral lifts and to the statics of the framework (Γ, p).

A stress $w \colon \Gamma_1 \to \mathbb{R}$ on a framework (Γ, p) is called a *self-stress* if it resolves the zero load:

$$\sum_{j \in \Gamma_0} w_{ij}(p_j - p_i) = 0 \text{ for all } i \in \Gamma_0. \tag{7}$$

Theorem 5.3. *Let (Γ, p) be a framework in $\mathbb{R}^2$ with a planar 3-connected graph Γ and such that for every face $\alpha \in \Gamma_2$ the points $\{p_i \mid i \in \alpha\}$ are not collinear. Then the following conditions are equivalent*:

1) *the framework has a self-stress that is non-zero on all edges*;
2) *the framework has a reciprocal diagram*;
3) *the framework has a vertical polyhedral lift*;
4) *the framework has a radial polyhedral lift.*

Proof. 1) $\Longrightarrow$ 2) From a self-stress w construct a reciprocal diagram (Γ^*, m) in the following recursive way. Take any face α_0 and define $m_{\alpha_0} \in \mathbb{R}^2$ arbitrarily. If for some face α the point m_α is already defined, then for every β adjacent to α put

$$m_\beta = m_\alpha + w_{ij} J(p_j - p_i),$$

where $J: \mathbb{R}^2 \to \mathbb{R}^2$ is the rotation by the angle $\frac{\pi}{2}$, ij is the edge dual to $\alpha\beta$, and the pair $(\alpha\beta, ij)$ is consistently oriented. In order to show that this gives a well-defined map $m: \Gamma_2 \to \mathbb{R}^2$, we need to check that the sum $\sum_{ij} w_{ij} J(p_j - p_i)$ vanishes along every closed path in the graph Γ^*. Viewed as a simplicial chain, every closed path is a sum of paths around vertices. The sum around a vertex vanishes due to (7). By construction, $m_\beta - m_\alpha \perp p_j - p_i$ and $m_\alpha \neq m_\beta$ for α and β adjacent in Γ^*, thus (Γ^*, m) is a reciprocal diagram for (Γ, p).

2) $\Longrightarrow$ 1) Let $(\alpha\beta, ij)$ be a consistently oriented dual pair. Since $m_\beta - m_\alpha \perp p_j - p_i$, there is $w_{ij} \in \mathbb{R}$ such that $m_\beta - m_\alpha = w_{ij} J(p_j - p_i)$. The map $w: \Gamma_1 \to \mathbb{R}$ thus constructed never vanishes and satisfies (7).

3) $\Longrightarrow$ 2) Given a polyhedral lift of (Γ, p), let $M_\alpha \subset \mathbb{R}^3$ be the plane to which the face α is lifted. Since M_α is not vertical, it is the graph of a linear function $f_\alpha: \mathbb{R}^2 \to \mathbb{R}$. Put $m_\alpha = \operatorname{grad} f_\alpha$. For every dual pair $(\alpha\beta, ij)$ we have

$$\tilde{p}_i, \tilde{p}_j \in M_\alpha \cap M_\beta.$$

This implies that the linear function $f_\alpha - f_\beta$ vanishes along the line through p_i and p_j, hence

$$m_\alpha - m_\beta = \operatorname{grad}(f_\alpha - f_\beta) \perp p_i - p_j.$$

2) $\Longrightarrow$ 3) Given a reciprocal diagram (Γ^*, m), construct a polyhedral lift recursively. Take any α_0 and let $f_{\alpha_0}: \mathbb{R}^2 \to \mathbb{R}$ be any linear function with $\operatorname{grad} f_{\alpha_0} = m_{\alpha_0}$. If f_α is defined for some α, then define f_β for every β adjacent to α by requiring

$$\operatorname{grad} f_\beta = m_\beta, \quad f_\beta - f_\alpha = 0 \text{ on the line } p_i p_j,$$

where ij is the edge dual to $\alpha\beta$. These conditions are consistent due to $m_\beta - m_\alpha \perp p_j - p_i$. In order to check that the recursion is well-defined, it suffices to show that if we start with some f_α and apply the recursion around the vertex $i \in \alpha$, then the new f_α will be the same as the old one. This is indeed the case because by construction all f_β with $i \in \beta$ take the same value at p_i. A polyhedral lift of (Γ, p) is obtained by putting $\tilde{p}_i = f_\alpha(p_i)$ for any $\alpha \ni i$.

3) $\iff$ 4) Consider $\mathbb{R}^3$ as an affine chart of $\mathbb{RP}^3$. There is a projective transformation $\Phi\colon \mathbb{RP}^3 \to \mathbb{RP}^3$ that restricts to the identity on $\mathbb{R}^2 \subset \mathbb{R}^3$ and sends the point a to the point at infinity that corresponds to the pencil of lines perpendicular to $\mathbb{R}^2$. (This transformation exchanges the plane at infinity with the plane through a parallel to $\mathbb{R}^2$.) We have $\mathrm{pr}_a = \mathrm{pr}_\perp \circ \Phi$. Therefore if $\tilde{p}$ is a radial polyhedral lift of p, then $\Phi \circ \tilde{p}$ is a vertical lift of p. Conversely, if $\tilde{p}$ is a vertical lift such that $\tilde{p}_i$ does not lie on the plane through a parallel to $\mathbb{R}^2$, then $\Phi^{-1} \circ \tilde{p}$ is a radial lift. Any orthogonal lift can be shifted in the direction orthogonal to $\mathbb{R}^2$ so that its vertices do not lie on the plane through a parallel to $\mathbb{R}^2$. Therefore the existence of an orthogonal lift is equivalent to the existence of a radial lift. ☐

Example 5.4. The Maxwell–Cremona correspondence allows to prove the rigidity criterium for the framework from Example 2.12. The lines a, b, c are concurrent if and only if the framework has a vertical lift, see Figure 11.

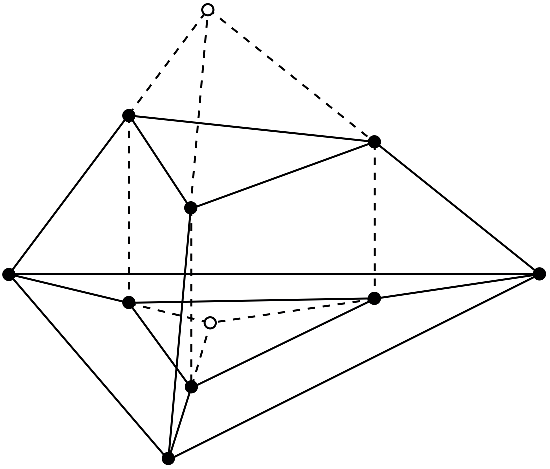


Figure 11. A vertical lift of the framework from Example 2.12

Remark 5.5. The spaces of self-stresses, reciprocal diagrams, and polyhedral lifts have natural linear structures. The correspondences described in the proof of Theorem 5.3 are linear, see also [11].

Every graph Γ has a geometric realization $|\Gamma|$: assign to the vertices points in $\mathbb{R}^3$ in general position, and to the edges the segments between those points. A map $\Gamma_0 \to \mathbb{R}^2$ can be extended to a map $|\Gamma| \to \mathbb{R}^2$ by affine interpolation. We call this the *rectilinear extension*. If the rectilinear extension is an embedding, then every face of Γ (viewed as a cycle of edges) becomes a polygon. In this case there is one face that is the union of all the other faces; we call it the *exterior face* (the term comes from the identification of $\mathbb{R}^2$ with a punctured sphere). The edges of the exterior face are called *boundary edges*, all of the other edges are called *interior edges*.

Theorem 5.6. *Let* (Γ, p) *be a framework in* $\mathbb{R}^2$ *with a planar* 3*-connected graph* Γ *and such that the rectilinear extension of* p *to* $|\Gamma|$ *provides an embedding of* Γ *into* $\mathbb{R}^2$ *with convex faces. Then the following conditions are equivalent*:

1) *the framework has a self-stress that is positive on all interior edges and negative on all boundary edges*;

2) *the framework has a reciprocal diagram such that for every consistently oriented dual pair* $(\alpha\beta, ij)$ *the pair of vectors* $(p_j - p_i, m_\beta - m_\alpha)$ *is positively oriented if* ij *is an interior edge and negatively oriented if* ij *is a boundary edge*;

3) *the framework has a vertical polyhedral lift to a convex polytope*;

4) *the framework has a radial polyhedral lift to a convex polytope.*

Proof. It suffices to show that the constructions in the proof of Theorem 5.3 respect the above properties.

1) $\iff$ 2) Since a self-stress is related to a reciprocal diagram by the formula $m_\beta - m_\alpha = w_{ij} J(p_j - p_i)$, the pair $(p_j - p_i, m_\beta - m_\alpha)$ is positively oriented if and only if $w_{ij} > 0$.

2) $\iff$ 3) Since $m_\beta - m_\alpha = \operatorname{grad}(f_\beta - f_\alpha)$, the pairs $(p_j - p_i, m_\beta - m_\alpha)$ for all interior edges ij are positively oriented if and only if the piecewise linear function over the union of the interior faces defined by $f(x) = f_\alpha(x)$ for $x \in \alpha$ is convex. The graph of this function together with the lift of the exterior face (that covers the union of the interior faces) form a convex polytope.

3) $\iff$ 4) The projective image of a convex polytope (provided no point is sent to infinity) is a convex polytope. The orthogonal lift can be made disjoint from the plane that is sent to infinity by shifting in the vertical direction. $\square$

Remark 5.7. By adding a linear function to an orthogonal polyhedral lift we can achieve that the exterior face stays in $\mathbb{R}^2$. A convex polytope of this kind is called a convex cap. An example is given on Figure 11.

Remark 5.8. The only self-intersections of the reciprocal diagram from Theorem 5.6 involve the edges $m_{\alpha_0} m_\beta$, where α_0 is the exterior face of Γ (and there is no way to get rid of all self-intersections unless Γ is the graph of the tetrahedron). The reciprocal diagram can be represented without self-intersections by replacing every edge $m_{\alpha_0} m_\beta$ with a ray running from m_β in the direction opposite to m_{α_0}. Complexes of this sort are called *spider webs* in [53].

Non-crossing frameworks with non-crossing reciprocals (and thus with some non-convex faces) are studied in the article [40].

Remark 5.9. The Dirichlet tesselation of a finite point set and the corresponding Voronoi diagram are a special case of a framework and a reciprocal diagram of the type described in Theorem 5.6. The vertical lift is given by $\tilde{p}_i = (p_i, \|p_i\|^2)$. The Voronoi diagram represents the reciprocal in the form of a spider web as described in the previous remark. A generalization of Dirichlet tesselations and Voronoi diagrams are weighted Delaunay tesselations and power diagrams. One of the definitions of a weighted Delaunay tesselation is a tesselation that possesses a vertical lift to a convex polytope. Thus one can add a fifth equivalent condition to Theorem 5.6: the framework is a weighted Delaunay tesselation. For details see [3].

In [53] the spider webs were related to planar sections of spatial Delaunay tesselations.

Remark 5.10. Not every convex tesselation and even not every triangulation of a convex polygon has a convex polyhedral lift, see [13, Chapter 7.1] for the "mother of all counterexamples." Those that do are called coherent or regular triangulations (more generally, tesselations). There is a generalization to higher dimensions, see [13].

5.3 Maxwell–Cremona correspondence in spherical geometry

Definition 5.11. Let (Γ, p) be a framework in $\mathbb{S}^2$ with a planar 3-connected graph Γ. A *weak reciprocal diagram* for (Γ, p) is a framework (Γ^*, m) in $\mathbb{S}^2$ such that

1) for every dual pair $(\alpha\beta, ij)$ the geodesics $p_i p_j$ and $m_\alpha m_\beta$ are perpendicular;

2) for every incident pair (α, i) the distance between m_α and p_i is different from $\frac{\pi}{2}$.

A *strong reciprocal diagram* is defined in the same way except that condition 2) is replaced by

2′) for every incident pair (α, i) the distance between m_α and p_i is less than $\frac{\pi}{2}$.

The reciprocity conditions can be rewritten as

$$\langle m_\alpha, p_i\rangle\langle m_\beta, p_j\rangle - \langle m_\alpha, p_j\rangle\langle m_\beta, p_i\rangle = 0 \tag{8}$$

$$\langle m_\alpha, p_i\rangle \neq 0 \tag{9}$$

$$\langle m_\alpha, p_i\rangle > 0 \tag{9$'$}$$

The left hand side in (8) equals $\langle m_\alpha \times m_\beta, p_i \times p_j\rangle$.

Definition 5.12. Let (Γ, p) be a framework in $\mathbb{S}^2$ with a planar 3-connected graph Γ and such that for every face $\alpha \in \Gamma_2$ the points $\{p_i \mid i \in \alpha\}$ are not collinear (that is, do not lie on a great circle). A *weak polyhedral lift* of (Γ, p) is a map $\tilde{p}\colon \Gamma_0 \to \mathbb{R}^3$ such that

1) $\tilde{p}_i = a_i p_i$ for every $i \in \Gamma_0$, where $a_i \neq 0$;

2) for every face $\alpha \in \Gamma_2$ the points $\{\tilde{p}_i \mid i \in \alpha\}$ are coplanar;

3) the planes of the adjacent faces differ from each other.

A *strong polyhedral lift* is defined similarly but with $a_i > 0$ in condition 1).

Theorem 5.13. *Let (Γ, p) be a framework in $\mathbb{S}^2$ with a planar 3-connected graph Γ and such that for every face $\alpha \in \Gamma_2$ the points $\{p_i \mid i \in \alpha\}$ are not collinear. Then the following conditions are equivalent:*

1) *the framework has a self-stress that is non-zero on all edges;*

2) *the framework has a weak reciprocal diagram;*

3) *the framework has a weak polyhedral lift.*

Proof. 1) $\Longrightarrow$ 3) By Lemma 3.7, a self-stress w gives rise to a map $\lambda\colon \Gamma_1 \to \mathbb{R}$ such that

$$\sum_{j\in\Gamma_0} \lambda_{ij} p_j \parallel p_i \quad \text{for all } i \in \Gamma_0. \tag{10}$$

Pick an $\alpha_0 \in \Gamma_2$ and define $\tilde{m}_{\alpha_0} \in \mathbb{R}^3$ arbitrarily. Define $\tilde{m}\colon \Gamma_2 \to \mathbb{R}^3$ recursively: if $\tilde{m}_\alpha$ is already defined, then for every β adjacent to α put

$$\tilde{m}_\beta = \tilde{m}_\alpha + \lambda_{ij}(p_i \times p_j),$$

where $(\alpha\beta, ij)$ is a consistently oriented dual pair. Equation (10) implies that the closing condition around every vertex i holds:

$$\sum_j \lambda_{ij}(p_i \times p_j) = 0.$$

Thus we have a well-defined map $\tilde{m}\colon \Gamma_2 \to \mathbb{R}^3$ with

$$\tilde{m}_\beta - \tilde{m}_\alpha \parallel p_i \times p_j$$

for any dual pair $(\alpha\beta, ij)$. In particular, $\tilde{m}_\beta - \tilde{m}_\alpha \perp p_i$, which implies that for every i there is $c_i \in \mathbb{R}$ such that

$$\langle \tilde{m}_\alpha, p_i \rangle = c_i$$

for all α incident to i. For a generic initial choice of $\tilde{m}_{\alpha_0}$ we have $c_i \neq 0$ for all i. If we put $\tilde{p}_i = \frac{p_i}{c_i}$, then we have

$$\langle \tilde{m}_\alpha, \tilde{p}_i \rangle = 1$$

for every incident pair (α, i). It follows that for every $\alpha \in \Gamma_2$ the points $\{\tilde{p}_i \mid i \in \alpha\}$ are coplanar and span a plane orthogonal to the vector $\tilde{m}_\alpha$. Due to $\lambda_{ij} \neq 0$ for every edge ij the planes of adjacent faces are different, thus we have constructed a weak polyhedral lift of (Γ, p).

3) $\Longrightarrow$ 2) Let $M_\alpha \subset \mathbb{R}^3$ be the plane containing the points $\{\tilde{p}_i \mid i \in \alpha\}$. Since the points $\{p_i \mid i \in \alpha\}$ are not collinear, the plane M_α does not pass through the origin. Thus it has an equation of the form

$$M_\alpha = \{x \in \mathbb{R}^3 \mid \langle \tilde{m}_\alpha, x \rangle = 1\}$$

for some $\tilde{m}_\alpha \in \mathbb{R}^3$. In particular, for any dual pair $(\alpha\beta, ij)$ we have

$$\langle \tilde{m}_\beta - \tilde{m}_\alpha, \tilde{p}_i \rangle = \langle \tilde{m}_\beta - \tilde{m}_\alpha, \tilde{p}_j \rangle = 0.$$

Hence the vector $\tilde{m}_\beta - \tilde{m}_\alpha$, and with it the plane spanned by $\tilde{m}_\alpha$ and $\tilde{m}_\beta$, is perpendicular to the plane spanned by p_i and p_j. If we put $m_\alpha = \frac{\tilde{m}_\alpha}{\|\tilde{m}_\alpha\|}$, then the geodesic $m_\alpha m_\beta$ is perpendicular to the geodesic $p_i p_j$. Since $\langle \tilde{m}_\alpha, \tilde{p}_i \rangle = 1$, we have $\langle m_\alpha, p_i \rangle \neq 0$. Thus (Γ^*, m) is a weak reciprocal diagram to (Γ, p).

2) $\Longrightarrow$ 3) Let (Γ^*, m) be a weak reciprocal diagram for (Γ, p). We construct lifts $\tilde{m}$ and $\tilde{p}$ such that

$$\langle \tilde{m}_\alpha, \tilde{p}_i \rangle = 1 \tag{11}$$

for every incident pair (α, i). The construction is recursive.

Pick $\alpha_0 \in \Gamma_2$ and lift m_{α_0} arbitrarily. Due to (9), for every $i \in \alpha_0$ there is a lift $\tilde{p}_i$ of p_i such that $\langle \tilde{m}_{\alpha_0}, \tilde{p}_i \rangle = 1$. If $\tilde{m}_\alpha$ is already defined, and β is adjacent to α, then let ij be the edge dual to $\alpha\beta$. First determine the lift $\tilde{p}_i$ from the condition (11), and then determine the lift $\tilde{m}_\beta$ from the same condition with β in place of α. Note that if we use p_j instead of p_i, then the result will be the same: due to the reciprocity conditions (8) and (9) we have

$$\langle \tilde{m}_\alpha, \tilde{p}_i \rangle = \langle m_\alpha, \tilde{p}_j \rangle \implies \langle \tilde{m}_\beta, \tilde{p}_i \rangle = \langle \tilde{m}_\beta, \tilde{p}_j \rangle.$$

This recursive procedure leads to well-defined lifts $\tilde{m}$ and $\tilde{p}$: going around a vertex i does not change the value of $\tilde{m}_\alpha$ because both the initial and the final values satisfy (11).

3) $\Longrightarrow$ 1) Let $\tilde{m}\colon \Gamma_2 \to \mathbb{R}^3$ be the map constructed during the proof of the implication 3) $\Longrightarrow$ 2). As it was shown, for every dual pair $(\alpha\beta, ij)$ the non-zero vector $\tilde{m}_\beta - \tilde{m}_\alpha$ is perpendicular to p_i and p_j. Thus we have a map $\lambda\colon \Gamma_1 \to \mathbb{R}$ such that

$$\tilde{m}_\beta - \tilde{m}_\alpha = \lambda_{ij}\, p_i \times p_j .$$

To determine the sign of λ_{ij}, we order the vertices so that the pair $(\alpha\beta, ij)$ is consistently oriented. Summing around a vertex i of Γ we obtain

$$p_i \times \sum_{j \in \Gamma_0} \lambda_{ij}\, p_j = 0.$$

Hence $\sum_{j \in \Gamma_0} \lambda_{ij}\, p_j \parallel p_i$ and by Lemma 3.7 the map λ gives rise to a non-zero self-stress on (Γ, p). ☐

We don't know what conditions on a framework and the stress guarantee the existence of a strong reciprocal diagram. At least it is necessary that the vertices of every face are contained in an open hemisphere. The next theorem shows that strong reciprocal diagrams correspond to strong polyhedral lifts.

Theorem 5.14. *Let* (Γ, p) *be a framework in* $\mathbb{S}^2$ *as in Theorem 5.13. Then the following conditions are equivalent*:

1) *the framework has a strong reciprocal diagram*;

2) *the framework has a strong polyhedral lift.*

Proof. In the proof of 3) $\Longrightarrow$ 2) in Theorem 5.13, note that for a strong lift $\tilde{p}$ the equation $\langle \tilde{m}_\alpha, \tilde{p}_i \rangle = 1$ implies $\langle m_\alpha, p_i \rangle > 0$, so that the reciprocal diagram constructed from a strong lift is strong itself.

In the proof of 2) $\Longrightarrow$ 3) in Theorem 5.13, lift m_{α_0} strongly (that is scale it by a positive factor). Condition (11) implies that all p_i with $i \in \alpha_0$ are also lifted strongly. The recursion propagates the strong lift to all m_β and p_j. ☐

As in the Euclidean case (see the paragraph before Theorem 5.6), a spherical framework defines a geodesic extension, that is, a map $|\Gamma| \to \mathbb{S}^2$ that sends every edge to an arc of a great circle. A geodesic extension is called a convex embedding of Γ if it is an embedding and every face is a convex spherical polygon.

Theorem 5.15. *Let* (Γ, p) *be a framework in* $\mathbb{S}^2$ *with a planar* 3*-connected graph* Γ *and such that its geodesic extension is a convex embedding. Then the following conditions are equivalent*:

1) *the framework has a self-stress that is positive on all edges*;
2) *the framework has a strong reciprocal diagram that embeds* Γ^* *into* $\mathbb{S}^2$ *with convex faces*;
3) *the framework has a strong lift to a convex polyhedron.*

Proof. 1) $\Longrightarrow$ 3) In the proof of the corresponding implication in Theorem 5.13 we have $\lambda_{ij} > 0$ for all edges ij. This implies that as we go around a vertex i, the vertices $\tilde{m}_\alpha$ for all α adjacent to i form a convex polygon. The union of these polygons is the boundary of a convex polyhedron that contains 0 in its interior. Its polar dual is a strong lift of (Γ, p).

3) $\Longrightarrow$ 2) A convex polyhedron that is a strong lift of (Γ, p) contains 0 in the interior. Thus its polar dual is also a convex polyhedron. The projection of the 1-skeleton of the dual is a strong reciprocal diagram with convex faces.

2) $\Longrightarrow$ 1) In a strong reciprocal diagram with convex faces the geodesics $m_\alpha m_\beta$ and $p_i p_j$ that correspond to a dual pair are consistently oriented. When we lift such a diagram as in the proof of 2) $\Longrightarrow$ 3) $\Longrightarrow$ 1) in Theorem 5.13, we obtain real numbers $\lambda_{ij} > 0$ that provide a positive self-stress on (Γ, p). ☐

The latter version of the spherical Maxwell–Cremona correspondence was described in [34].

Remark 5.16. As in the Euclidean case, not every convex tesselation of the sphere has a convex polyhedral lift. The corresponding theory predates the theory of regular triangulations in the Euclidean space and was developed by Shephard [48] and McMullen [35]. See also [17].

5.4 Maxwell–Cremona correspondence in hyperbolic geometry

Let (Γ, p) be a framework in $\mathbb{H}^2$ with a planar 3-connected graph Γ. A *reciprocal diagram* is a framework (Γ^*, m) in $\mathbb{H}^2$ such that for every dual pair $(\alpha\beta, ij)$ the geodesics $m_\alpha m_\beta$ and $p_i p_j$ are perpendicular. In terms of the Minkowski scalar product this means

$$\langle m_\alpha, p_i \rangle \langle m_\beta, p_j \rangle - \langle m_\alpha, p_j \rangle \langle m_\beta, p_i \rangle = 0.$$

Remark 5.17. The above criterion of orthogonality of $m_\alpha m_\beta$ and $p_i p_j$ as well as its spherical analog (8) can be reformulated as follows. Diagonals in a spherical or hyperbolic quadrilateral with side lengths a, b, c, d in this cyclic order are orthogonal if and only if

$$\cos_{\mathbb{X}} a \cos_{\mathbb{X}} c = \cos_{\mathbb{X}} b \cos_{\mathbb{X}} d.$$

The diagonals of a Euclidean quadrilateral are orthogonal if and only if $a^2 + c^2 = b^2 + d^2$.

Definition 5.18. Let (Γ, p) be a framework in $\mathbb{H}^2$ with a planar 3-connected graph Γ and such that for every face $\alpha \in \Gamma_2$ the points $\{p_i \mid i \in \alpha\}$ are not collinear. A *polyhedral lift* of (Γ, p) is a map $\tilde{p}\colon \Gamma_0 \to \mathbb{R}^3$ such that

1) $\tilde{p}_i = a_i p_i$ for every $i \in \Gamma_0$, where $a_i > 0$;
2) for every face $\alpha \in \Gamma_2$ the points $\{\tilde{p}_i \mid i \in \alpha\}$ are contained in a space-like plane;
3) the planes of the adjacent faces differ from each other.

Theorem 5.19. *Let (Γ, p) be a framework in $\mathbb{H}^2$ with a planar 3-connected graph Γ and such that for every face $\alpha \in \Gamma_2$ the points $\{p_i \mid i \in \alpha\}$ are not collinear. Then the following conditions are equivalent:*

1) *the framework has a self-stress that is non-zero on all edges;*
2) *the framework has a reciprocal diagram;*
3) *the framework has a polyhedral lift.*

Proof. 1) $\Longrightarrow$ 3) Proceed as in the proof of Theorem 5.13 to construct a map $\tilde{m}\colon \Gamma_2 \to \mathbb{R}^3$ such that

$$\tilde{m}_\beta - \tilde{m}_\alpha \parallel p_i \times p_j$$

(with the Minkowski cross-product) for every dual pair $(\alpha\beta, ij)$. By changing the position of $\tilde{m}_{\alpha_0}$ and scaling down the self-stress w we can achieve that all $\tilde{m}_\alpha$ belong to the upper half of the light cone. Then the planes $\langle \tilde{m}_\alpha, x\rangle = -1$ bound a polyhedron with space-like faces that is a polyhedral lift of (Γ, p).

3) $\Longrightarrow$ 2) Similarly to the proof of Theorem 5.13, let $\langle \tilde{m}_\alpha, x\rangle = -1$ be an equation of the plane containing the points $\{\tilde{p}_i \mid i \in \alpha\}$. Since these planes are space-like, $\tilde{m}_\alpha$ are time-like, and since $\tilde{p}_i$ belongs to the upper half of the light cone, $\tilde{m}_\alpha$ also does. Hence (Γ^*, m) is a reciprocal diagram in $\mathbb{H}^2$.

2) $\implies$ 3) The proof is the same as in Theorem 5.13, we lift (Γ, p) and (Γ^*, m) recursively and at the same time.

3) $\implies$ 1) Also the same as in Theorem 5.13, but with the Minkowski cross-product instead of the Euclidean. ☐

For a framework (Γ, p) in $\mathbb{H}^2$ the geodesic extension $|\Gamma| \to \mathbb{H}^2$ is an analog of the rectilinear extension in the Euclidean case: an edge ij of Γ is mapped to the geodesic segment $p_i p_j$. If the geodesic extension is an embedding, then we define the interior and exterior faces and the interior and boundary edges as in the Euclidean case, see the paragraph before Theorem 5.6.

For a consistently oriented dual pair $(\alpha\beta, ij)$ we say that the lines $p_i p_j$ and $m_\alpha m_\beta$ are consistently oriented if the directed line $m_\alpha m_\beta$ is obtained from the directed line $p_i p_j$ through rotation by $\frac{\pi}{2}$ around their intersection point.

Theorem 5.20. *Let* (Γ, p) *be a framework in* $\mathbb{H}^2$ *with a planar* 3*-connected graph* Γ *and such that the geodesic extension of* p *to* $|\Gamma|$ *provides an embedding of* Γ *into* $\mathbb{H}^2$ *with convex faces. Then the following conditions are equivalent*:

1) *the framework has a self-stress that is positive on all interior edges and negative on all boundary edges*;

2) *the framework has a reciprocal diagram such that for every consistently oriented dual pair* $(\alpha\beta, ij)$ *the lines* $p_i p_j$ *and* $m_\alpha m_\beta$ *are consistently oriented if* ij *is an interior edge and non-consistently oriented if* ij *is a boundary edge*;

3) *the framework has a polyhedral lift to a convex polytope in the Minkowski space.*

Proof. The proof consists in checking that the constructions in the proof of Theorem 5.19 respect the above properties. ☐

Remark 5.21. A variant of the Maxwell–Cremona theorem for hyperbolic frameworks uses an orthogonal polyhedral lift to the co-Minkowski space instead of a radial polyhedral lift to the Minkowski space described above. For details on the co-Minkowski space see [16].

Remark 5.22. If we allow the faces of the polyhedral lift to be time-like or light-like, then the vertices of the corresponding reciprocal diagram become de Sitter or ideal. Since the reciprocity is a symmetric notion, it is natural to allow de Sitter and ideal positions for the vertices of the framework as well. This puts us into the more general context of hyperbolic-de Sitter frameworks or point-line-horocycle frameworks, see Section 2.3.

References

[1] A. D. Alexandrov, *Convex polyhedra,* translated from the 1950 Russian edition by N. S. Dairbekov, S. S. Kutateladze, and A. B. Sossinsky, with comments and bibliography by V. A. Zalgaller and appendices by L. A. Shor and Yu. A. Volkov, Springer Monographs in Mathematics, Springer-Verlag, Berlin, 2005.
MR 2127379 Zbl 1067.52011 ☞ 199

[2] L. Asimow and B. Roth, "The rigidity of graphs. II," *J. Math. Anal. Appl.* 68 (1979), no. 1, 171–190.
MR 531431 Zbl 0441.05046 ☞ 197

[3] F. Aurenhammer, R. Klein, and D.-T. Lee, *Voronoi diagrams and Delaunay triangulations,* World Scientific, Hackensack, N.J., 2013.
MR 3186045 Zbl 1295.52001 ☞ 221

[4] W. Blaschke, "Über affine Geometrie XXVI: Wackelige Achtflache," *Math. Zeitschr.* 6 (1920), 85–93.
JFM 47.0681.05 ☞ 202

[5] E. D. Bolker and B. Roth, "When is a bipartite graph a rigid framework?," *Pacific J. Math.* 90 (1980), no. 1, 27–44.
MR MR599317 Zbl 0407.05078 ☞ 201

[6] R. Bricard, "Mémoire sur la théorie de l'octaèdre articulé," *Journ. de Math.* (5) 3 (1897), 113–148.
☞ 202

[7] A.-L. Cauchy, "Sur les polygones et polyèdres, second mémoire," *Journal de l'Ecole Polytechnique* 19 (1813), 87–98.
☞ 199

[8] R. Connelly, "The rigidity of certain cabled frameworks and the second-order rigidity of arbitrarily triangulated convex surfaces," *Adv. in Math.* 37 (1980), no. 3, 272–299.
MR MR591730 Zbl 0446.51012 ☞ 197

[9] R. Connelly, "Rigidity," in *Handbook of convex geometry* (P. M. Gruber and J. M. Wills, eds.), Vol. A, North-Holland, Amsterdam, 1993, 223–271.
MR MR1242981 Zbl 0788.52001 ☞ 192

[10] R. Connelly, "Generic global rigidity," *Discrete Comput. Geom.* 33 (2005), no. 4, 549–563.
MR 2132290 Zbl 1072.52016 ☞ 198

[11] H. Crapo and W. Whiteley, "Spaces of stresses, projections and parallel drawings for spherical polyhedra," *Beiträge Algebra Geom.* 35 (1994), no. 2, 259–281.
MR 1312669 Zbl 0819.52018 ☞ 219

[12] G. Darboux, *Leçons sur la théorie générale des surfaces,* III, IV, Lignes géodésiques et courbure géodésique, Paramètres différentiels, Déformation des surfaces, Déformation infiniment petite et représentation sphérique, reprint of the 1894 original (III) and the 1896 original (IV), Les Grands Classiques Gauthier-Villars, Cours de Géométrie de la Faculté des Sciences, Éditions Jacques Gabay, Sceaux, 1993.
MR 1365962 ☞ 192

[13] J. A. De Loera, J. Rambau, and F. Santos, *Triangulations,* Structures for algorithms and applications, Algorithms and Computation in Mathematics 25, Springer-Verlag, Berlin, 2010.
MR 2743368 Zbl 1207.52002 ☞ 221

[14] M. Dehn, "Über die Starrheit konvexer Polyeder," *Math. Ann.* 77 (1916), 466–473.
JFM 46.1115.01 Zbl 46.1115.01 ☞ 199

[15] Y. Eftekhari, B. Jackson, A. Nixon, B. Schulze, S.-i. Tanigawa, and W. Whiteley, "Point-hyperplane frameworks, slider joints, and rigidity preserving transformations," *Journal of Combinatorial Theory*, Series B, to appear. ☞ 198

[16] F. Fillastre and A. Seppi, "Spherical, hyperbolic, and other projective geometries: convexity, duality, transitions," in *Eighteen essays in non-Euclidean geometry* (V. Alberge and A. Papadopoulos, eds.), European Mathematical Society (EMS), Zürich, 2019, Chapter 16, 321–409.
☞ 211, 227

[17] F. Fillastre and I. Izmestiev, "Shapes of polyhedra, mixed volumes, and hyperbolic geometry," *Mathematika* 63 (2017), no. 1, 124–183.
MR 3610008 Zbl 1367.52010 ☞ 225

[18] A. A. Gaifullin, "Flexible cross-polytopes in spaces of constant curvature," *Tr. Mat. Inst. Steklova* 286 (2014), 88–128, in Russian. English translation in *Proc. Steklov Inst. Math.* 286 (2014), no. 1, 77–113.
MR 3482593 Zbl 1320.52015 ☞ 202

[19] H. Gluck, “Almost all simply connected closed surfaces are rigid,” in *Geometric topology* (L. C. Glaser and T. B. Rushing, eds.) Proceedings of the Geometric Topology Conference held at Park City, Utah, February 19–22, 1974, Lecture Notes in Mathematics 438. Springer-Verlag, Berlin etc., 1975, 225–239.
MR MR0400239 Zbl 0315.50002 ☞ 197

[20] I. Ivanova-Karatopraklieva, P. E. Markov, and I. Kh. Sabitov, “Bending of surfaces. III,” *Fundam. Prikl. Mat.* 12 (2006), no. 1, 3–56, in Russian. English translation in *J. Math. Sci.* (*N.Y.*) 149 (2008), no. 1, 861–895.
MR 2249679 Zbl 1149.53301 ☞ 192

[21] I. Ivanova-Karatopraklieva and I. Kh. Sabitov, “Deformation of surfaces. I,” in *Проблемы геометрии* (*Problems in geometry,* N. M. Ostianu, ed.), Vol. 23, Итоги Науки и Техники, Akad. Nauk SSSR, Vsesoyuz. Inst. Nauchn. i Tekhn. Inform., Moscow, 1991, 131–184, 187, in Russian. English translation in *J. Math. Sci.* 70 (1994), no. 2, 1685–1716.
MR 1152588 ☞ 192

[22] I. Ivanova-Karatopraklieva and I. Kh. Sabitov, “Bending of surfaces. II,” Geometry, 1. *J. Math. Sci.* 74 (1995), no. 3, 997–1043.
MR 1330961 Zbl 0861.53002 ☞ 192

[23] I. Izmestiev, “Spherical and hyperbolic conics,” in *Eighteen essays in non-Euclidean geometry* (V. Alberge and A. Papadopoulos, eds.), European Mathematical Society (EMS), Zürich, 2019, Chapter 15, 263–320.
☞ 202

[24] I. Izmestiev, “Projective background of the infinitesimal rigidity of frameworks,” *Geom. Dedicata* 140 (2009), 183–203.
MR 2504741 Zbl 1169.52010 ☞ 208, 209

[25] I. Izmestiev, “Examples of infinitesimally flexible 3-dimensional hyperbolic cone-manifolds,” *J. Math. Soc. Japan* 63 (2011), no. 2, 581–598.
MR 2793111 Zbl 1221.57029 ☞ 202

[26] I. Izmestiev, “Classification of flexible Kokotsakis polyhedra with quadrangular base,” *Int. Math. Res. Not. IMRN* 2017, no. 3, 715–808.
MR 3658150 ☞ 204

[27] B. Jackson and J. C. Owen, “A characterisation of the generic rigidity of 2-dimensional point–line frameworks,” *J. Combin. Theory Ser. B* 119 (2016), 96–121.
MR 3486339 Zbl 1338.52021 ☞ 198

[28] B. Jessen, “Orthogonal icosahedra,” *Nordisk Mat. Tidskr* 15 (1967), 90–96.
MR MR0226494 Zbl 0156.41001 ☞ 203

[29] A. Kokotsakis, “Über bewegliche Polyeder,” *Math. Ann.* 107 (1933), no. 1, 627–647.
MR 1512819 ☞ 204

[30] G. Laman, “On graphs and rigidity of plane skeletal structures,” *J. Engrg. Math.* 4 (1970), 331–340.
MR 0269535 Zbl 0213.51903 ☞ 198

[31] L, Lecornu, “Sur l’equilibre des surfaces flexibles et inextensibles,” *J. de l’Éc. Pol. Cah.* XLVIII (1880), 1-109.
JFM 16.0784.01 ☞ 194

[32] A.-M. Legendre, *Éléments de géométrie, avec des notes,* Firmin Didot, 1794 (an II), 321–334.
☞ 199

[33] H. Liebmann, “Ausnahmefachwerke und ihre Determinante,” *Münch. Ber.* 50 (1920), 197–227.
JFM 47.0710.02 ☞ 202, 210

[34] L. Lovász, “Steinitz representations of polyhedra and the Colin de Verdière number,” *J. Combin. Theory Ser. B* 82 (2001), no. 2, 223–236.
MR 1842113 Zbl 1023.05100 ☞ 225

[35] P. McMullen, “Representations of polytopes and polyhedral sets,” *Geometriae Dedicata* 2 (1973), 83–99.
MR 0326574 Zbl 0273.52006 ☞ 225

[36] A. F. Möbius, “Kann von zwei dreiseitigen Pyramiden eine jede in Bezug auf die andere um- und eingeschrieben zugleich heißen?,” *J. Reine Angew. Math.* 3 (1828), 273–278.
ERAM 003.0113cj ☞ 202

[37] A. F. Möbius, *Lehrbuch der Statik,* Georg Joachim Göschen, Leipzig, 1837.
☞ 192

[38] G. Nawratil, “Flexible octahedra in the projective extension of the Euclidean 3-space,” *J. Geom. Graph.* 14 (2010), no. 2, 147–169.
MR 2799365 Zbl 1235.52032 ☞ 202

[39] Y. Nishiyama, “Miura folding: applying origami to space exploration,” *Int. J. Pure Appl. Math.* 79 (2012), no. 2, 269–279.
Zbl 1254.51009 ☞ 204

[40] D. Orden, G. Rote, F. Santos, B. Servatius, H. Servatius, and W. Whiteley, “Non-crossing frameworks with non-crossing reciprocals,” *Discrete Comput. Geom.* 32 (2004), no. 4, 567–600.
MR 2096749 Zbl 1062.05047 ☞ 221

[41] A. V. Pogorelov, *Внешняя геометрия выпуклых поверхностей,* Izdat. “Nauka,” Moscow, 1969, in Russian. English translation, *Extrinsic geometry of convex surfaces,* Translated by Israel Program for Scientific Translations, Translations of Mathematical Monographs 35, American Mathematical Society, Providence, R.I., 1973.
MR 0244909 MR0346714 (transl.) Zbl 0311.53067 ☞ 194, 211

[42] L. Poinsot, *Éléments de statique,* Calixte-Volland, Paris, 1803 (an XII).
☞ 192

[43] W. J. M. Rankine, “On the application of barycentric perspective to the transformation of structures,” *Phil. Mag. Series* 4 (1863), no. 26, 387–388.
☞ 210

[44] I. Kh. Sabitov, “Local theory of bendings of surfaces,” in *Geometry* III, Theory of surfaces, a translation of Current problems in mathematics, Fundamental directions 48 (Russian), Akad. Nauk SSSR, Vsesoyuz. Inst. Nauchun. i Tekhn. Inform., Moscow, 1989, Translation by E. Primrose, Translation edited by Yu. D. Burago and V. A. Zalgaller, Encyclopaedia of Mathematical Sciences 48. Springer-Verlag, Berlin, 1992, 179–256.
MR 1039820 MR 1306736 (transl.) ☞ 192

[45] R. Sauer and H. Graf, “Über Flächenverbiegung in Analogie zur Verknickung offener Facettenflache,” *Math. Ann.* 105 (1931), no. 1, 499–535.
MR 1512724 JFM 57.0875.03 Zbl 0002.41101 ☞ 204

[46] R. Sauer, “Projektive Sätze in der Statik des starren Körpers,” *Math. Ann.* 110 (1935), no. 1, 464–472.
MR MR1512950 ☞ 194, 210

[47] E. Schönhardt, “Über die Zerlegung von Dreieckspolyedern in Tetraeder,” *Math. Ann.* 98 (1927), 309–312.
JFM 53.0576.01 ☞ 202

[48] G. C. Shephard, “Spherical complexes and radial projections of polytopes,” *Israel J. Math.* 9 (1971), 257–262.
MR 0282291 Zbl 0207.52202 ☞ 225

[49] H. Stachel, “Zur Einzigkeit der Bricardschen Oktaeder,” *J. Geom.* 28 (1987), no. 1, 41–56.
MR 878483 Zbl 0611.51011 ☞ 202

[50] E. Steinitz, “Polyeder und Raumeinteilungen,” in *Encyclopädie der mathematischen Wissenschaften,* Dritter Band: Geometrie, III.1.2., Heft 9, Kapitel 3 A B 12, 1922, 1–139.
☞ 216

[51] W. Whiteley, “Infinitesimal motions of a bipartite framework,” *Pacific J. Math.* 110 (1984), no. 1, 233–255. MR MR722753 Zbl 0476.70003 ☞ 201

[52] W. Whiteley, “Infinitesimally rigid polyhedra. I. Statics of frameworks,” *Trans. Amer. Math. Soc.* 285 (1984), no. 2, 431–465.
MR MR752486 Zbl 0518.52010 ☞ 199

[53] W. Whiteley, P. F. Ash, E. Bolker, and H. Crapo, “Convex polyhedra, Dirichlet tessellations, and spider webs,” in *Shaping space* (M. Senechal, ed.), Exploring polyhedra in nature, art, and the geometrical imagination, with the assistance of G. Fleck and S. Sherer, Springer, New York, 2013, 231–251.
MR 3077229 ☞ 221

[54] H. Whitney, “Congruent graphs and the connectivity of graphs,” *Amer. J. Math.* 54 (1932), no. 1, 150–168.
MR 1506881 JFM 58.0609.01 Zbl 0003.32804 ☞ 216

[55] G. M. Ziegler, *Lectures on polytopes,* Graduate Texts in Mathematics 152, Springer-Verlag, New York, 1995.
MR 1311028 Zbl 0823.52002 ☞ 216

PART II
Projective geometries

Chapter 13

Contributions to non-Euclidean geometry I [*]

Eduard Study

The author's intention is to edit a series of articles under a common title concerning algebraic and differential geometry in non-Euclidean space, a subject that he has been investigating for many years although not yet arriving at any other than sketchy publications on the topic.

In the first of the present three contributions,[1] a seemingly new property of the straight line will be discussed. The second article is independent from the first. It is entitled "the notions of left and right,"[2] but it also treats other fundamental notions of non-Euclidean geometry. It is meant to be an introduction to further articles, and therefore it is kept elementary. It may perhaps also serve as an introduction to the methods that the author used in his book *Geometrie der Dynamen* (Leipzig, 1903). The third note contains a simple application to helical surfaces in non-Euclidean space of positive curvature that are considered as geodesics in a four-dimensional manifold.

I. Straight line and point as extrema

It seems not to have been noticed so far that the theorem of the straight line as the shortest path between two points has two analoga in hyperbolic geometry, where the *point* remarkably does not appear as the shortest but as the *longest path between two lines*.

If we extend the point continuum of *Lobatchevsky's* geometry to projective geometry by adding the infinitely far, or so-called ideal points in the well-known way, then one can also state this claim that obviously needs to be made more precise, in correlative form: *the straight line or rather a piece of the straight line appears as the longest path between two points.*

---

[*] "Beitrage zur Nicht-Euklidischen Geometrie. I.", *Amer. J. Math.* 29 (1907), no. 2, 101–116. Translated by Annette A'Campo-Neuen.

[1] [Editor's note] In the present volume, only the first contribution is translated, whose subtitle is *Gerade und Punkt als Extreme*, translated, as below, by *Straight line and point as extrema*.

[2] We plan to show elsewhere that a part of the concept of the notions of left and right does not need the notion of measurement but that it can be considered in the real projective geometry of space.

We first consider the plane geometry of constant curvature -1, associated to the absolute polar system

$$(xy) = x_1y_1 - x_2y_2 - x_3y_3 = 0. \tag{1}$$

Then we can uniquely define the distance (that is here always considered as *positive* ≥ 0) between two points in the interior $\{(xx) > 0, (yy) > 0\}$ of the absolute cone[3] by the formula

$$(x, y) = \operatorname{arccosh} \sqrt{\frac{(xy)^2}{(xx)(yy)}}, \tag{2}$$

where we choose the positive value for the square root and the function arccosh. The same formula provides a real "distance" for the points x, y, even in the case when they lie outside of the absolute cone such that the straight line joining them meets the absolute cone. Moreover, also in this case for three points y, x, z on a straight line, where the one in the middle x, separates the other two, the following equation holds:

$$(y, z) = (y, x) + (x, z). \tag{3}$$

Now consider a triangle x, y, z of distinct points in the exterior such that each of the lines joining two of the points meets the absolute cone. Hence

$$\begin{aligned} (yz)^2 - (yy)(zz) \geq 0, \quad (xx) < 0, \\ (zx)^2 - (zz)(xx) \geq 0, \quad (yy) < 0, \\ (xy)^2 - (xx)(yy) \geq 0, \quad (zz) < 0. \end{aligned} \tag{4}$$

Then one can distinguish two families of such triangles. If

$$(yz)(zx)(xy) < 0, \tag{5a}$$

then the longest side of the triangle x, y, z is always greater than the sum of the other two, or at least equal to their sum. And equality holds if and only if the three points are collinear.

If however

$$(yz)(zx)(xy) > 0, \tag{5b}$$

then the longest side may be greater or smaller than the sum of the other two, or it may be equal to their sum. But in this case the three points are never collinear.

---

[3] [Translator's note] The author uses the term "absoluter Kegelschnitt," i.e. absolute conic section, for the zero set of the bilinear form. Indeed the lines through zero on the cone form a conic section in the projective plane.

Let us now assume that

$$(y,z) \geq (y,x), \quad (y,z) \geq (x,z), \tag{6}$$

and ask under what assumptions we can conclude that

$$(y,z) \geq (y,x) + (x,z). \tag{7a}$$

Obviously, we can rewrite inequality (7a) in this form:

$$\cosh((y,z) - (y,x)) \geq \cosh(x,z).$$

After solving by using the addition theorem for the hyperbolic cosine on the left-hand side, the inequality takes the following form, assuming positive values wherever a square root is taken:

$$\sqrt{(yz)^2}\,\sqrt{(xy)^2} + (yy)\sqrt{(xz)^2} \geq \sqrt{(yz)^2 - (yy)(zz)}\,\sqrt{(xy)^2 - (xx)(yy)}.$$

Here, by the assumptions (4) and (6), both sides are positive. Hence the inequality is equivalent to the one that we obtain by taking squares:

$$(xx)(yy)(zz) - 2\sqrt{(yz)^2\,(zx)^2\,(xy)^2} - (xx)(yz)^2 - (yy)(zx)^2 - (zz)(xy)^2 \geq 0.$$

On the other hand,

$$(xx)(yy)(zz) + 2(yz)(zx)(xy) - (xx)(yz)^2 - (yy)(zx)^2 - (zz)(xy)^2 \geq 0,$$

since the left-hand side is equal to the square of the determinant (xyz). Therefore, if the inequality (5a) holds, then the last two inequalities coincide and (7a) holds. They are both equalities if $(xyz) = 0$. However, we cannot conclude that conversely (7a) implies (5a).

Now let us assume that (5b) holds. Then we claim that any of these relations are possible

$$(y,z) \gtreqless (y,x) + (x,z), \tag{7b}$$

or as above (assuming (6) *analogously*) the following may occur

$$(xyz)^2 - 4(yz)(zx)(xy) \gtreqless 0.$$

In fact, one can verify this easily by examples or by the argument given below. In this case (xyz) never vanishes because then in (7b) we would have equality. Therefore the product $(yz)(zx)(xy)$ would also vanish. But this is impossible by (4).

In order to obtain a clear idea of the corresponding triangular figures we want to assume that the points y, z are given. It is easy to see that then the set of all points x satisfying the inequalities (5a) or (5b) is represented by the shaded regions in Figure 1.

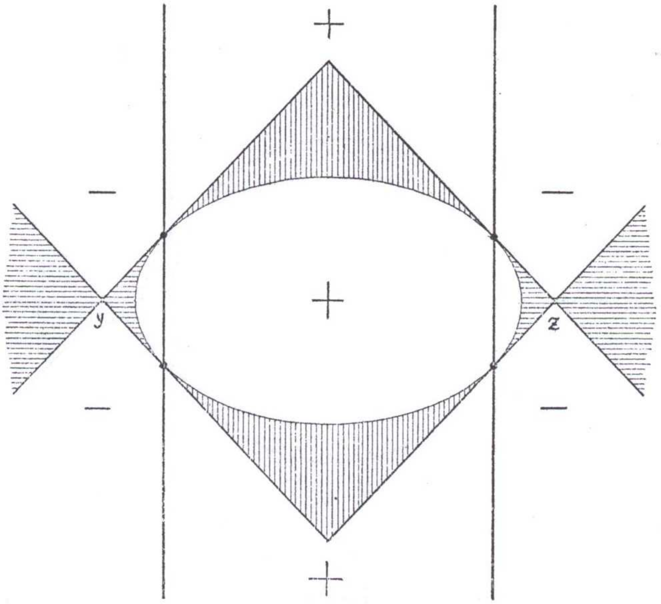


Fig. 1.

Vertical shading indicates that

$$(yz)(zx)(xy) > 0,$$

and horizontal shading indicates that

$$(yz)(zx)(xy) < 0.$$

Moreover, Figure 2 indicates with vertical and horizontal shading where

$$(y, z) - (y, x) - (x, z) \geq 0, \quad \text{or} \quad (y, z) - (y, x) - (x, z) \leq 0.$$

The figure also shows with fat lines the locus of all points x where

$$(y, z) - (y, x) - (x, z) = 0.$$

This locus consists of three completely separate curves. One of these curves is linear whereas the other two are parts of the curve of order 2

$$(yzx)^2 - 4(yz)(yx)(zx) = 0. \tag{8}$$

Firstly, this curve touches the polar lines of the points y, z, that is the lines $(yx) = 0$ and $(zx) = 0$, at their intersection points with $(yzx) = 0$. Now if we write down its equation in line coordinates using the notation

$$(ux) = x_1u_1 + u_2x_2 + u_3x_3, \quad (uv) = u_1v_1 - u_2v_2 - u_3v_3,$$

then after splitting off the non-vanishing factor $4(yz)\{(yz)^2 - (yy)(zz)\}$ from the quadratic covariant of the expression (8), one finds the equation

$$(yz)(uu) - (uy)(uz) = 0. \tag{9}$$

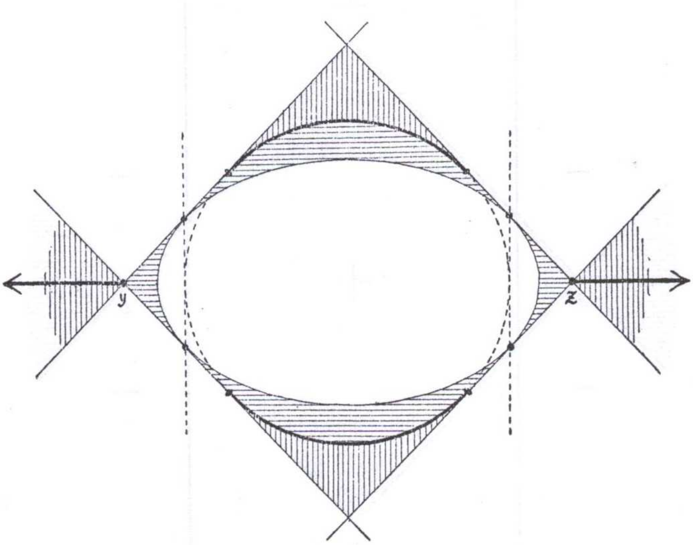


FIG. 2.

Hence one can see that the curve (8) touches the minimal lines through y, z, i.e. the tangents to the absolute cone going through y, z, of course in real points. Finally, through the two touching points of the curve (8) with the minimal lines through z there is the locus of $(y, x) = (y, z)$ which is not drawn in Figure 2, but could easily be added. The regions within the curves $(y, x) = (y, z)$ and $(z, x) = (y, z)$ belong to the inequalities $(y, z) < (y, x)$ and $(y, z) < (z, x)$, and are therefore horizontally shaded. The curve (8) however, separates the two kinds of shading wherever it lies *outside* of these regions. Now it is easy to see that the shadings are carried out correctly by letting x run through the boundary curves of the different regions.

For greater clarity, we add here a third figure—in collinear modification—showing for a given x, where the points y and z have to be if we require simultaneously

$$(yz)(zx)(xy) < 0 \quad \text{and} \quad (y, z) \geq (y, x) + (x, z). \tag{10}$$

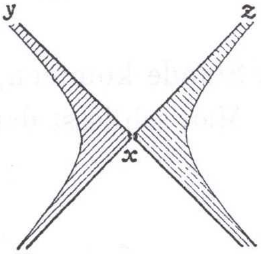


FIG. 3.

In this case, which will now be investigated further, we will say: “The point x separates the points y and z,” or “The point x lies between y and z.” If there is no minimal line among the lines joining the points x, y, z then the polar lines of the points x, y, z are in the relative position indicated in Figure 4a, whereas the triangles of the second family correspond to Figure 4b.

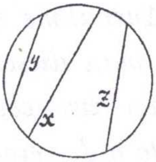


FIG. 4a.

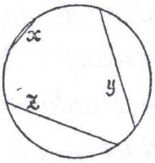


FIG. 4b.

In order to avoid lengthy arguments, for the following theorems we refer to intuition (see Figure 4a), or an obvious continuity consideration:

If the point x' lies between y and z, and the point x'' lies between (y and x' or) x' and z, then x'' also lies between y and z.

If two different points lie between y and z in such a way that the straight line joining them meets the absolute cone then all four points have a certain order y, x', x'', z or z, x'', x', y. More precisely, after suitable choice of notation, x' lies between y and x'', and x'' lies between x' and z, but not x'' between y and x', and not x' between x'' and z.

If, moreover, the point x lies between two such points x' and x'', then x also lies between y and z.[4]

---

[4] All this could be proved rigorously with purely algebraic means but as far as we can see the proof would not really be short. Unless it concerns fundamental theorems of deep theories, we may well permit ourselves to shorten the presentation if it is possible to fill the gaps easily if needed.

The two drawings in Figure 5, may sufficiently illustrate how those theorems are derived.

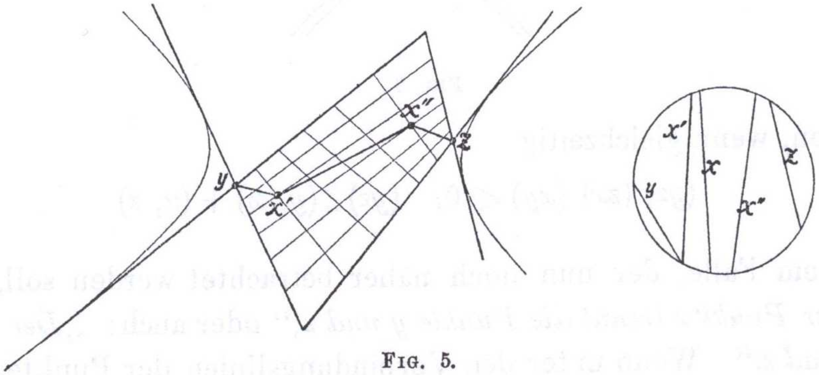


FIG. 5.

Now it is easy to conclude the following remarkable theorem:

Suppose that there is a given set of points outside of the absolute cone, forming a continuum, and such that each straight line joining any two different points of the set meets the absolute cone. Then every point of the set with at most two exceptions y_0, z_0, *has the following property: If this point* x *is removed from the set then the remaining points of the set form two again continuous subsets* (y) *and* (z) *such that* x *separates all the points* y *of the first set from all the points* z *of the second set.*

All the points of the set lying between two different points y, z *form together with* y, z *a closed, continuous curve whose points* x *can be assigned bijectively and continuously the values of a parameter* t *between* 0 *and* 1.

This curve has a well-determined length S *satisfying the inequalities*

$$0 \leq S \leq (y, z).$$

Moreover, if there are two exceptional points y_0, z_0 as mentioned above, then the entire set consists of these two points and a chain of curves joining them with a certain arc length and whose points different from y_0, z_0 all lie between y_0 and z_0. The points y, z referred to in the theorem always preserve the property of such limit points y_0, z_0 if one removes all those points of the set different from y, z that do *not* lie between y and z.

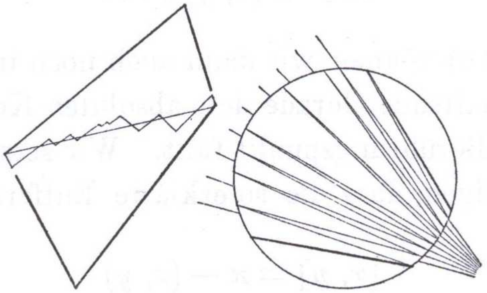


FIG. 6.

Such a chain of curves as described in the theorem will be called a *path* between y and z. Now we have the result announced in the beginning:

If y *and* z *are two different points outside of the absolute cone, whose joining straight line meets the absolute cone, then the straight line is the longest among all paths between* y *and* z.

Note these further useful remarks: *If the straight line joining the points* y, z *is a minimal line, i.e. if it touches the absolute cone, then this straight line is the unique path between* y *and* z. *In every other case, there are infinitely many paths between* y *and* z, *and arbitrarily close to any such path there are infinitely many other paths of length zero.*

Figure 6 illustrates how to construct a path that is arbitrarily close to the longest path, but itself is of length zero.

For the points of a straight line *not* intersecting the absolute cone, i.e. a line lying entirely in the ideal domain, we can define an elliptic measurement.[5] The distance of two points x, y on such a line will then only be determined up to sign and up to a multiple of a certain period, that we can assume to be π. Among the different values of this function, we choose the one with smallest absolute value, consider it as a positive value and denote it by (x, y). If moreover, we choose for all occurring roots their positive values, we obtain the equation

$$\cos(x, y) = \sqrt{\frac{(xy)^2}{(xx)(yy)}}, \tag{11}$$

and we always have

$$0 \leq \cos(x, y) \leq 1.$$

---

[5] [Translator's note] The author uses the term "Massbestimmung." It is not a metric however, since for the points in the exterior the triangle inequality does not always hold.

We can think of the same formula (11) as being true in the limiting case where the straight line containing x, y *touches* the absolute cone, but neither x nor y is the point of contact. We then set $(x, y) = 0$.

Now we want to show that the hereby defined distance (x, y) as well as its supplement

$$[x, y] = \pi - (x, y)$$

also has the property of being a *maximum* among all the lengths of certain "paths" between the two points, that still have to be defined.

Again we consider a triangle x, y, z of different ideal points but now such that the straight lines joining any two points shall *not meet or at most touch* the absolute cone:

$$\begin{aligned} (yz)^2 - (yy)(zz) &\leq 0, \quad (xx) < 0, \\ (zx)^2 - (zz)(xx) &\leq 0, \quad (yy) < 0, \\ (xy)^2 - (xx)(yy) &\leq 0, \quad (zz) < 0. \end{aligned} \tag{12}$$

Without loss of generality, we may assume that

$$\frac{\pi}{2} \geq (y, z) \geq (y, x) \geq (x, z) \geq 0. \tag{13}$$

The inequalities that we will have to assume consecutively

$$(y, z) \gtreqless (y, x) + (x, z) \tag{14}$$

may then be written in the following form:

$$\cos(x, z) - \cos(y, z)\cos(y, x) \gtreqless \sin(y, z) \cdot \sin(y, x).$$

Since here by assumption both sides are positive values, taking squares leads to equivalent inequalities

$$(xx)(yy)(zz) - 2\sqrt{(yz)^2(zx)^2(xy)^2} - (xx)(yz)^2 - (yy)(zx)^2 - (zz)(xy)^2 \gtreqless 0.$$

As before one can see that here only $>$ or equality is possible if

$$(yz)(zx)(xy) \leq 0, \tag{15}$$

whereas in the case

$$(yz)(zx)(xy) > 0 \tag{16}$$

one cannot draw a conclusion.

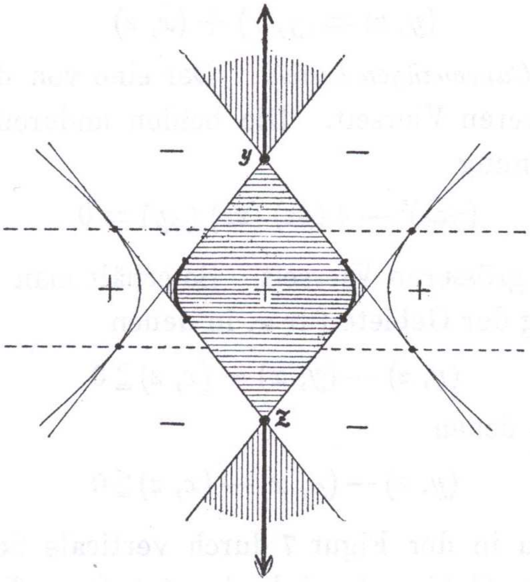


Fig. 7.

It will be sufficient if moreover we first exclude the easily treated limiting cases

$$(y,z)=\frac{\pi}{2}=[y,z]\quad\text{and}\quad(y,z)=0,[y,z]=\pi$$

and hence assume that

$$0<(y,z)<\frac{\pi}{2}<[y,z]<\pi.$$

Each of the two segments into which the straight line yz is decomposed by the two points y,z is now a diagonal of a simple quadrangle whose sides are pieces of minimal lines (compare Figure 7).

In the interior or on the boundary of these two quadrangles are all the points x satisfying the inequalities (12). Apart from that both quadrangles have different properties. By integrating the arc element along a straight line between y and z through the interior of one quadrangle one obtains the smaller value (y,z), and by integrating along a straight line through the other quadrangle one obtains the larger value $[y,z]$. The first “smaller" quadrangle as we call it for simplicity, is completely contained in the domain $(yz)(zx)(xy)<0$. The other “larger" quadrangle is decomposed by the absolute polar lines of the points y, z into three parts, two outer parts where also $(yz)(zx)(xy)<0$, and one inner part, where $(yz)(zx)(xy)>0$. Moreover, analogously to above one finds *that the locus of all points x with*

$$(y,z)=(y,x)+(x,z) \tag{17}$$

consists of three separate curves. The first one of these is linear and lies entirely in the smaller quadrangle. The other two lie on the irreducible curve of degree 2

$$(yzx)^2 - 4(yz)(zx)(xy) = 0 \tag{18}$$

and are contained entirely in the larger quadrangle. Thus one finally obtains insight into the distribution of the regions of x where

$$(y,z) - (y,x) - (x,z) \geq 0$$

holds and the other ones with

$$(y,z) - (y,x) - (x,z) \leq 0.$$

In Figure 7, the first ones are indicated by vertical shading and the second ones or more precisely the *one* second domain is indicated by horizontal shading.

Now as above for the hyperbolic measurement, we obtain a notion of *path* between y and z lying in the smaller quadrangle, and we find the theorem that among all such paths the straight line is the longest. We further claim that in the larger quadrangle the same relations hold. That is, we claim that in the triangle x, y, z the length of the side yz is not less than the sum of the lengths of the other two sides if x is contained in the larger quadrangle and if all three lengths are defined via integration of the arc element along a straight line such that the path of integration does not leave the larger quadrangle.

One can see without difficulty that for the proof of this claim one has to distinguish three cases.

1) The point x lies in the middle part of the larger quadrangle (or on its boundary). Then $(yz)(yx)(zx) \geq 0$. The three lengths are $[y,z]$, (y,x) and (x,z). The claim is

$$[y,z] \geq (y,x) + (x,z).$$

2) and 3) The point x lies in one of the outer parts, e.g. in the part cut off by the polar line of the point y, or such that $(y,x) \geq (x,z)$. Then $(yz)(yx)(zx) \leq 0$, the three lengths are $[y,z]$, $[y,x]$, (x,z), and the claim is

$$[y,z] \geq [y,x] + (x,z) \quad \text{or} \quad (y,x) \geq (y,z) + (z,x).$$

In all these cases the proof can be carried out in the same manner as above.

Now we will state the result in a shorter form than previously and we will essentially only point out what is different from the previous theorem.

We call the used kind of measurement *pseudo-elliptic* since it coincides with a usual elliptic one in constant positive curvature $(+1)$, but differs substantially from it by the existence of paths of length zero. Using this terminology we can then say:

If y, z are two different points whose straight line joining them does not intersect the absolute cone, then there are two families of direct paths between y and z. Namely, there are two simple quadrangles bounded by minimal lines having y and z as opposite vertices. In the interior or on the boundary of each of these quadrangles there are the paths of the first or the second family respectively. Among all paths of the same family, using pseudo-elliptic measurement, the straight line appears as the longest.

Each direct path is a piece of a curve of a certain arc length whose secants either do not meet or at most touch the absolute cone. By any of its points x different from y and z, it is decomposed into two paths with the same property. If in the limiting case that we first excluded, y and z remain ideal points such that the straight line yz touches the cone then there is exactly *one* “proper” and *one* “improper” direct path between y and z. Both are linear. The first one lies entirely outside of the absolute cone; it has length zero. The length of the second path is not properly defined, but maybe set $= \pi$.

A piece of a curve can have the property that around each of its points one can draw a region where the piece of the curve is a direct path between any two of its points y', z', even though it may not have this property as a whole. For such pieces of curves, joining two given points y, z, the last theorem does not hold. On the contrary, one can obviously join two arbitrary ideal points—even those whose joining straight line intersects the absolute cone in separate points—by such curves whose length exceeds an arbitrary given number. The length of the direct path between the points y, z, if such a path exists, is hence in a much more restricted sense a maximum than the path length considered with respect to hyperbolic measurement.

Since the introduction of the so-called ideal points in Lobatchevsky’s geometry is not really necessary, one might prefer a different formulation of the presented theorems avoiding that notion. Of course the translation from one language into the other can be carried out easily. For example the first result that we discussed can be rephrased as follows:

If among three straight lines y, x, z, that are pairwise either parallel or ultraparallel, the middle one separates the other two, then the length of the common normal between y and z is greater or equal to the sum of the lengths of the normals between y, x and between x, z.

Equality occurs if all three lines have a common perpendicular, or in the limiting case, if they are all parallel. If y, z are ultraparallel, the "length of the common normal between y and z" means the length of the *shortest* joining line between a point of y and a point of z, and in the limiting case of being parallel, length zero. Everything else follows immediately: The ideal point of intersection of two ultraparallel lines y, z, or its polar line, the common perpendicular of y, z, considered as the locus of the lines perpendicular to it, now appears as the *longest path* between these two lines.

Similarly, one can translate our second theorem with its corollaries. The fundamental fact in this case turns out not to be different from the well-known theorem on angles of a triangle, that can be expressed as follows:

In each ordinary triangle one of the outer angles is greater or equal to the sum of the two inner angles that occur at the side opposite to the outer angle. And equality holds if and only if the three sides go through one common point.

It is less intuitive to work with lines and their angles rather than with points and their distances. In the first case one does not have as in the second case a guiding principle in the analogous situation in Euclidean geometry. And that is the reason why up to now, one has not given this well-known result the turn on to which we focus here (if indeed the matter might be new). But if we present the straight line here as a however relative *maximum*, then this is certainly a fact that should not be missing in a complete system of non-Euclidean geometry. More generally, it may well be permitted to mention that as it seems to us, the development of non-Euclidean geometry is still very incomplete even in very elementary aspects. For example Lobatchevsky's trigonometry is certainly a very important part but always only a part: triangles with all or some ultraparallel sides, in particular the hexagons with only right angles corresponding to the first kind, have also to be considered in a system that claims some kind of completeness. The theory available at present still lacks finer perfection, the investigation of the corresponding groups, the connection to invariant theory etc., even though the tools for such investigations have long since been provided for.[6]

---

[6] The methods and formulae available in the literature are completely sufficient for those purposes. See S. W. Jacobsthal, article "Trigonometrie" in *Encyclopädie der elementaren Geometrie* (Leipzig, 1905), and E. Study, "Sphärische Trigonometrie" etc. (*Abh. d. Sächs.,* G. D. W. Leipzig, 1893), and E. Study, *Geometrie der Dynamen,* Leipzig 1903, pp. 209–212.

Now asking for an extension of the deduced results to arbitrary dimension, we find a remarkable difference among the sketched little theories.

Our first theorem on the straight line as longest path can be generalized to an arbitrary dimension, the second theorem only holds for the plane.

Namely, let the absolute polar system be given by

$$(xy) = x_1y_1 - x_2y_2 - \cdots - x_ny_n = 0.$$

Now first consider two points y, z outside of the absolute cone $(xx) = 0$ such that the straight line joining them meets the cone. In every plane through the line yz, we obtain the triangle figures that we have already investigated. So for hyperbolic measurement, instead of the quadrangle that we discussed above, we obtain a piece of the space of dimension $(n-1)$, bounded by pieces of two conoids emanating from y and z

$$(yx)^2 - (yy)(xx) = 0, \quad (zx)^2 - (zz)(xx) = 0.$$

Now "paths" in this piece of space have to be defined basically as before, but they do not have to lie in planes. The straight path is the longest among all such paths.

However, if the line yz does not meet the absolute cone then there are three different kinds of planar paths between y, z. Firstly, a plane through y, z might permeate the absolute cone in a real curve. For pseudo-elliptic measurement, (y, z) or $[y, z]$ appears as the maximum among all suitable piecewise linear paths. Secondly, the plane might touch the absolute cone. It is easy to see that in this case there are infinitely many equally long paths between both points. Finally, the plane may not intersect the absolute cone at all. In this case (y, z) is a minimum. *Hence the length of the straight linear path in this case is not extremal.*

But this obviously does not exclude the possibility that on a manifold of lower dimension with such a measurement, the problem of shortest (or longest) lines may get a clear content. For example in the eight-dimensional space, there is a four-dimensional manifold M_4 with the following strange property for suitably chosen pseudo-elliptic measurement:

The length of the shortest path measured on M_4 itself between two points y, z in M_4 equals the length of the straight linear path between these two points.[7]

We only want to sketch other generalizations of the matters discussed. Similar extremal properties as the ones considered here, can be found for geodesics belonging to an arc element whose square is defined by a quadratic differential form in the case that this form for line elements yields a hyperbolic measurement. Moreover, in some parts of our previous considerations, one can replace the cone by an arbitrary oval. Somewhat related to these ideas is for example the following easily proved but maybe still not completely uninteresting theorem:

---

[7] E. Study, "Kürzeste Wege im komplexen Gebiet," *Math. Ann.* 60 (1905), p. 361.

Consider a set of points in the finite part of the Euclidean plane forming a closed continuum and having the property that all secants between different points of the set have an angle φ *with a given straight line satisfying an inequality of the form*

$$|\tan(\varphi)| \leq M.$$

Then this point set is a rectifiable curve joining certain points y, z. *If* D *is the distance of the projection of the points* y, z *to the given line, then the arc length of this curve is at most equal to*

$$|D\sqrt{1+M^2}|.$$

Chapter 14

Notes on Eduard Study's paper "Contributions to non-Euclidean geometry I"

Annette A'Campo-Neuen and Athanase Papadopoulos

Contents

1. An outline of Study's paper 253
2. Notes on Study's paper 257
3. A note on Eduard Study 259

References 261

The paper "Contributions to non-Euclidean geometry I" [10] by Eduard Study, published in 1907, contains one of the first attempts to study the geometry of the exterior of the hyperbolic plane. In particular, the author defines a distance between points in the exterior of the disc model of this space sitting in the projective plane, and for triples of points in this exterior space, he discusses at length the various cases where the triangle inequality is satisfied and those where a reverse triangle inequality is satisfied. The setting is that of projective geometry. In the present note, we provide an outline of Study's paper, together with some historical and mathematical comments.

1 An outline of Study's paper

1.1 A model of the hyperbolic plane

In the three-dimensional vector space $\mathbb{R}^3$ equipped with a basis in which points are represented by coordinates as $x = (x_1, x_2, x_3)$, the author starts with the quadratic form of index two defined by

$$(x, x) = x_1^2 - x_2^2 - x_3^2. \tag{1}$$

To this quadratic form is associated an *absolute cone* (the isotropic cone of the quadratic form) whose equation is

$$(x, x) = 0.$$

The *interior* of the absolute cone is the set of points x satisfying

$$(x, x) > 0.$$

A *distance function* (x, y) is defined on the interior of the absolute cone by the formula

$$(x, y) = \operatorname{arccosh} \sqrt{\frac{(xy)^2}{(xx)(yy)}}. \tag{2}$$

Considering (x_1, x_2, x_3) as the homogeneous coordinates of a point x in the 2-dimensional real projective space and using the homogeneity of formula (2), the author gets a distance function on the quotient of the interior of the absolute cone by the positive reals, and this distance turns out to be a model of the hyperbolic plane. The goal of the paper is to extend this distance function to the points that are outside of this hyperbolic plane.

1.2 The exterior of the hyperbolic plane

The same formula (2) is used to define a distance function on the exterior of the absolute cone in $\mathbb{R}^3$. In the quotient projective plane, this will provide a distance function on the exterior of the hyperbolic plane, making this space a Lorentzian surface which is not Riemannian. The distance function is first defined on *some* pairs of points (x, y), namely, those for which the (Euclidean) straight line joining them meets the absolute cone.

This distance function, as in the case of the above model of the hyperbolic plane, is “projective” in the sense that it satisfies the property that if three points x, y, z lie in that order on a Euclidean segment, we have

$$(x, z) = (x, y) + (y, z).$$

In other words, Euclidean lines are geodesics for this distance function. For points x, y, z satisfying this equation, y is said to lie *between* x and z.

Study considers *triangles* in this Lorentzian surface, that is, triples of points such that for each pair among them, the Euclidean line joining them meets the absolute cone. Certain triples of points satisfy the triangle inequality, whereas others satisfy the reverse triangle inequality. More precisely, Study proves the following:

Proposition 1.1. *For a triangle x, y, z in the exterior hyperbolic geometry,*

(1) *if $(xy)(yz)(zx) < 0$ then the longest side of the triangle is at least equal to the sum of the other two*;

(2) *if $(xy)(yz)(zx) > 0$ then the longest side of the triangle may be greater, equal or smaller than the sum of the other two. Furthermore, in this case, the three points can never be collinear.*

We note that distance functions satisfying the reverse triangle inequality are studied in the context of the geometrical setting of relativity theory, and the reverse triangle inequality is called in this setting the *time inequality*; cf. Busemann's paper [2].

Study provides detailed descriptions of all the situations given in Proposition 1.1 (Figures 1, 2, and 3 of his paper).

1.3 The space of lines in the hyperbolic plane

There is a polarity theory between the hyperbolic plane and its exterior where the points in the exterior represent lines in the hyperbolic plane, and vice versa, the points in hyperbolic plane represent lines in the exterior of that plane. Using this polarity, Study interprets the results stated above in terms of distances between lines in the hyperbolic space, and he provides pictorial representations (Figures 4 and 5 of his paper).

1.4 Geodesics

Study proves the following theorem which he describes as being "remarkable," regarding geodesics for the distance function defined outside the absolute cone. The theorem has two parts:

Theorem 1.2. *The following holds.*

(1) *Consider a set of points outside the absolute cone forming a continuum*[1] *and such that each straight line joining any two points of the set meets the absolute cone. Then every point x of the set, with at most two exceptions x_0 and y_0, has the following property: the complement of the point x in the set is a union of two continua such that x separates all points y from the first set and points z from the second set.*

---

[1] The terminology is not uniform. In the classical terminology, a *continuum* can be a closed connected subset of a topological space, or a compact connected metric space. In the present situation the two terminologies coincide. (There are other notions of continua.)

(2) *Within the set of points described in* (1), *the set of points that lie between any two distinct points x and y form, together with x and y a continuous curve homeomorphic to the segment* $[0, 1]$ *whose length is at most equal to the distance between x and y.*

A curve such as the one described in (2) of the proposition is called a path between x and y. The fact that its length is less than or equal to the distance between x and y follows from the observation that the entire path must lie within the region where the reverse triangle inequality holds.

Study provides a description of the exceptional set of points x_0, y_0 mentioned in the first part of Theorem 1.2.

A second theorem is stated:

Theorem 1.3. *If x and y are two distinct points outside the absolute cone such that the straight line joining them intersects the absolute cone, then this straight line is the longest among all the paths joining x and y.*

This gives an explanation of the following sentence at the beginning of the paper: "It seems not to have been noticed so far that the theorem of the straight line as the shortest path between two points has two analoga in hyperbolic geometry, where the point remarkably does not appear as the shortest but as the longest path between two lines."

A remark is made for the case where the straight line joining x and y is tangent to the absolute cone. In this case, this straight line is the unique path joining x and y. In all the other cases, there are infinitely many paths between these two points and arbitrarily close to each such path, there are infinitely many others whose lengths are zero.

1.5 A measure on a line not intersecting the absolute cone

Study then introduces a distance (x, y) between pairs of points (x, y) such that the line that joins them does not intersect the absolute cone. The distance is defined up to a sign and up to a period. Such a distance function is termed a "pseudo-elliptic measure." Taking the sign for the distance between two points to be positive, the measure is given by the formula:

$$\cos(x, y) = \sqrt{\frac{(xy)^2}{(xx)(yy)}}. \tag{3}$$

In the limiting case where the given line touches the absolute cone, with neither x nor y being the point of contact, one sets, by continuity, $(x, y) = 0$.

This distance function between two points is equal to the length of the line segment joining them and its value is a maximum among all lengths of a certain collection of paths which Study defines between two points and which he studied in a certain detail.

The term "pseudo-elliptic" measure used for this distance function is chosen because of the analogy with the distance function on the sphere. In fact, taking as (xy) the Euclidean inner product in $\mathbb{R}^3$ instead of the one associated to the quadratic form (1), formula (3) is a consequence of the usual cosine formula in $\mathbb{R}^3$ applied to the triangle Oxy where O is the origin of the coordinates. It also gives the formula for the cosine of the distance between the points x and y on the sphere. (Recall that in the usual model of the sphere, centered at the origin of $\mathbb{R}^3$, the distance between two points is defined as the angle between the rays starting at the origin and passing through these two points.)

At the end of the paper, Study discusses higher dimensional generalizations. He declares that Theorem 1.2 admits a generalization whereas Theorem 1.3 is essentially planar.

He also declares that non-Euclidean geometry has to be further developed, for instance, completing Lobachevsky's formulae by including triangles or right angled hexagons with some edges being ultraparallel (that is, having some vertices at infinity), including also group theory in this domain, making connections with invariant theory, etc.

2 Notes on Study's paper

Study uses the terms "distance" and "measure" between points. Both terms were used by Klein in his fundamental paper [7] in which he defines distance functions associated to conics that give rise to non-Euclidean metrics (see also [1]). One should recall here that the terminology and the axioms for a metric space as we understand them today were still not standardized, and we cannot expect Study to talk about a "metric space" in this sense. As a matter of fact, the axioms for a metric space were formulated by Fréchet in his 1906 paper [5] (which was also his doctoral dissertation), and by the time Study wrote his paper, this set of axioms was still far from being universally adopted as a definition of a metric.

In conclusion, there was no way in which Study could discuss the question of whether his distances and measures give rise to genuine metric spaces in the sense we intend them today.

In his paper "A sixth memoir upon quantics" [3], Cayley defined a distance on the sphere by the formula using cosines in Study's paper (formula (3) above). Cayley's context, like Study's, is that of projective geometry, but with an absolute cone associated to an imaginary conic and given by the equation

$$x_1^2 + x_2^2 + x_3^3 = 0.$$

In his paper [7] (Section 11), Klein used Cayley's formula to define a distance he called a "measure." He then transformed this formula to obtain the known distance formula in terms of the cross-ratio.

Cayley later expressed his satisfaction regarding Klein's formula. In his comments on his paper [3] that he included in Volume II of his *Collected mathematical papers* ([4], p. 604), he writes:

> In his first paper, Klein substitutes, for my $\cos^{-1}$ expression for the distance between two points, a logarithmic one; viz. in linear geometry if the two fixed points are A, B then the assumed definition for the distance of any two points P, Q is
>
> $$\text{dist.}(PQ) = s \log \frac{AP.BQ}{AQ.BP};$$
>
> this is a great improvement, for we at once see that the fundamental relation,
>
> $$\text{dist.}(PQ) + \text{dist.}(QR) = \text{dist.}(PR),$$
>
> is satisfied.

The formulae to which Cayley refers are contained in his paper [3] p. 584.

Measuring distances of points in the exterior of the hyperbolic space, with the idea that these points represent hyperplanes of the space itself, and making trigonometric formulae for this exterior geometry is now a well established topic. It is addressed for instance in §2.6 of Thurston's notes on the geometry and topology of 3-manifolds, see [14].

The reader will have noticed that Study declares in his introduction that the results in his article are part of a subject he has been investigating for many years "although not yet arriving at anything else than sketchy publications on the topic." It was not unusual, among nineteenth century German mathematicians, to publish papers in which they give the important ideas without the details of proofs. Sometimes, theorems were stated without the precise hypotheses that describe the conditions of their validity. Klein's "Ueber die sogenannte Nicht-Euklidische Geometrie" [7] is an example of a paper written in this style.

It would be useful to rewrite Study's paper with more details, making explicit the relation with the de Sitter metric and other works, in particular Poincaré's quadratic geometries ([8] and the prologue of the present volume) and the distance functions related to the geometric theory of relativity, e.g. Busemann's timelike metrics [2].

3 A note on Eduard Study

Christian Hugo Eduard Study (1862–1930) was mainly interested in invariant theory. He had a very personal approach to mathematics. Despite the fact that he studied at university (first, biology, for one year, in Jena, then mathematics in Strasbourg, Leipzig, and Munich), he considered himself as self-taught. Apparently, he did not like to listen to lectures but he used to read lecture notes and rethink the material in his own way (cf. [6]). In a letter to Klein, talking about Christoffel's lectures in Strasbourg, Study writes: "All these I can obtain in the form of good notes that I actually prefer to the lectures themselves. Because here I am free and independent, I can study whenever I want, I can skip well-known material or make my own selection and thereby proceed much faster than if I would have to sit still on the seat during one semester" (our translation from [6], p. 50). On the other hand, it appears from the biography in [6] that Study appreciated the inspiration he got from his teachers, in particular Klein, and admired Lie and Grassmann.

In 1885, Study obtained a position at the University of Leipzig where Klein and Hilbert were working, and, starting from 1894, he taught at the University of Bonn. It is believed that Study had a crucial influence on Felix Hausdorff, whose career was stagnant, partly because of his extreme modesty, encouraging him and leading him towards important research problems. Among Study's subjects of interest, we mention spherical trigonometry; cf. his paper [11] and his book [12]. He interacted and corresponded with Lie, Engel, Gordan, Severi and several other mathematicians, but he remained a solitary person.

In a letter to Hurwitz, quoted by Constance Reid in her book [9] (p. 25), Hilbert writes: "[Study] is a strange person, and almost completely at opposite poles from my nature and, as I think I can judge, from yours too. Dr. Study approves, or rather he knows, only one field of mathematics and that's the theory of invariants, very exclusively the symbolic theory of invariants. Everything else is unmethodical 'fooling around' …He condemns for this reason all other mathematicians; even in his own field he considers himself to be the only authority, at times attacking all the other mathematicians of the symbolic theory of invariants in the most aggressive fashion.

He is one who condemns everything he doesn't know whereas, for example, my nature is such I am most impressed by just that which I don't yet know." To this harsh comment by Hilbert, one may add the more positive words by Klein on Study's dissertation: "Dr. Study is of independent, sensitive nature. If he moreover learns to submit his subjective impulses to the requirements of given conditions then I hope his habilitation will contribute substantially to support mathematical studies at our university" (our translation from [6] p. 57).

Study had an unusual—sometimes awkward—style which we perceive in the paper that concerns us here. Let us quote Study from an answer he published in the Bulletin of the AMS [13] to a critique made by M. Bôcher on one of his papers: "What I am concerned with is merely the inference which the author leaves to his readers. This inference would appear to be that I had been thoughtless, or careless, or regardless enough to publish my results in an embryonic state of development. It will inevitably be understood that all my theorems are 'true in general' only, and consequently incorrect. [...] My fault (if fault it is) thus consisted in supposing notions widely known (and accepted by competent judges) which in reality perhaps were not so. Why I held such an opinion is easily explained. The ideas in question are very simple. From the standpoint of Klein's Program of 1872 they appear almost as matters of course. To me they have been familiar since about 1884, the time when I was a young student under Professor Klein, who in later years was Professor Bôcher's teacher also, and to whom also Professor Bôcher, as he says himself, is indebted for the same ideas. It would seem to follow from this that the public I had in mind was likely to exist [...]"

Let us also quote a citation from a report on an article StudyStudy, Eduard wrote while he was a Ph.D. student in Munich, in which he gives a response to a problem raised by A. Brill concerning spacial curves of degree 4, and for which he won a prize: "In his introduction, the author aims at developing the theory of invariants on a geometrical basis and hereby arrives at a—though complicated—geometric interpretation of the process of symbolic calculation. Even though, concerning the gaps in the proofs given in the introduction, one cannot completely agree with the author's opinion about the importance of such investigations, nevertheless one has to praise and appreciate the independent spirit and the talent with which the author works for his self-chosen goal" (our translation from [6], p. 46)

Finally, let us mention that Gian-Carlo Rota who edited a reprint of Study's book on ternary forms in 1984 wrote about this book: "the flamboyant German style and the scattered remarks, often sparkling and original, make the reading enjoyable."

The Ph.D. dissertation by Y. Hartwig [6], entitled *Eduard Study* (*1862–1930*)*—ein mathematischer Mephistopheles im geometrischen Gärtchen*, contains other interesting material on Study, namely correspondence with Klein, Engel and others as well as reports on his dissertation, habilitation, etc.

References

[1] N. A'Campo and A. Papadopoulos, "On Klein's 'So-called non-Euclidean geometry,'," in *Sophus Lie and Felix Klein*: *The Erlangen program and its impact in mathematics and in physics* (L. Ji and A. Papadopoulos, eds.), IRMA Lectures in Mathematics and Theoretical Physics 23, European Mathematical Society (EMS), Zürich, 2015. 91–136.
MR 3362214 ☞ 257

[2] H. Busemann, ”Timelike spaces,” *Dissertationes Math. Rozprawy Mat.* 53 (1967) 52 pp. Reprinted in *Selected works* (A. Papadopoulos, ed.), Vol. I, Springer Verlag, Cham, 2018, 741–786.
MR 0220238 Zbl 0156.43201 ☞ 255, 259

[3] A. Cayley, "A sixth memoir upon quantics," *Philos. Trans. R. Soc. Lond.* 149 (1859), 61–90. Reprinted in *The collected mathematical papers of Arthur Cayley* (A. R. Forsyth, ed.), Vol. II, Cambridge University Press, Cambridge, 1888, 561–592.
☞ 258

[4] A. Cayley, *The collected mathematical papers of Arthur Cayley* (A. R. Forsyth, ed.), Vol. II, Cambridge University Press, Cambridge, 1888.
JFM 20.0024.02 ☞ 258

[5] M. Fréchet, "Sur quelques points de calcul fonctionnel," *Rend. Circ. Mat. Palermo* 22 (1906), 1-74.
☞ 257

[6] Y. Hartwig, *Eduard Study* (*1862–1930*)*—ein mathematischer Mephistopheles im geometrischen Gärtchen,* Ph.D. thesis, Johannes Gutenberg-Universität, Mainz, 2005.
☞ 259, 260

[7] F. Klein, "Ueber die sogenannte Nicht-Euklidische Geometrie," *Clebsch Ann.* IV, (1871), 573-625; *Gött. Nachr.* 17 (1871), 419-433 (shorter version). French translation, "Sur la géométrie dite non euclidienne" (J. Hoüel, translator), *Darboux Bull.* II (1871), 341–351. Reprint *Gesammelte mathematische Abhandlungen* (R. Fricke and A. Ostrowski, eds.), Springer, Berlin, 1921, 254–305. English translation in J. Stillwell, *Sources of hyperbolic geometry,* History of Mathematics, 10. American Mathematical Society, Providence, R.I., and London Mathematical Society, London, 1996, 69–111.
JFM 03.0231.02 ☞ 257, 258

[8] H. Poincaré, "Sur les hypothèses fondamentales de la géométrie," *Bull. Soc. Math. France* 15 (1887), 203–216.
JFM 19.0512.01 ☞ 259

[9] C. Reid, *Hilbert,* With an appreciation of Hilbert's mathematical work by H. Weyl, Springer-Verlag, Berlin etc., 1970.
MR 0270884 Zbl 0192.32601 ☞ 259

[10] E. Study, "Beiträge zur nichteuklidischen Geometrie I., II., III.," *Amer. J. Math.* 29 (1907), pp. 101–167. English translation of the first paper by A. A'Campo-Neuen, "Contributions to non-Euclidean geometry I," in *Eighteen essays in non-Euclidean geometry* (V. Alberge and A. Papadopoulos, eds.), European Mathematical Society (EMS), Zürich, 2019, Chapter 13, 237–251.
JFM 38.0503.03 ☞ 253

[11] E. Study, "Über die sphärische Trigonometrie," Jahresbericht DMV Bd. 2, 1891.
☞ 259

[12] E. Study, *Sphärische Trigonometrie, orthogonale Substitutionen, und elliptische Functionen: Eine Analytisch-Geometrische Untersuchung*. Teubner, Leipzig 1893.
JFM 25.0798.02 ☞ 259

[13] E. Study, "Professor Bôcher's views concerning the geometry of inversion," *Bull. Amer. Math. Soc.* 22 (1915), no. 1, 38–40.
MR 1559705 JFM 45.1345.01 JFM 45.0940.05 ☞ 260

[14] W. P. Thurston, *The geometry and topology of three-manifolds,* mimeographed notes, Princeton University, Princeton, N.J., 1979.
`http://library.msri.org/books/gt3m/`
☞ 258

Chapter 15

Spherical and hyperbolic conics

Ivan Izmestiev

Contents

1. Introduction 263
2. Quadratic forms, conics, and pencils 265
3. Spherical conics 276
4. Theorems about spherical conics 282
5. Hyperbolic conics 293
6. Theorems about hyperbolic conics 304

References 317

1 Introduction

Most textbooks on classical geometry contain a chapter about conics. There are many well-known Euclidean, affine, and projective properties of conics. For broader modern presentations we recommend the corresponding chapters in the textbook of Berger [2] and two recent books [1] and [13] devoted exclusively to conics. For a detailed survey of the results and history before the 20th century, see the encyclopedia articles [6] and [7].

At the same time, one rarely speaks about spherical and hyperbolic conics. However, they should not be seen as something exotic. Recall that a *projective conic* is essentially a quadratic cone in $\mathbb{R}^3$ (that is, the zero set of a homogeneous quadratic polynomial). An *affine conic* is the intersection of a quadratic cone with an affine plane. A *Euclidean conic* is an affine conic in the presence of a Euclidean metric in the plane.

Now, a *spherical conic* is a projective conic in the presence of the canonical metric on the space of rays, the distance between two rays being the angle between them. In other words, a spherical conic is the intersection of a quadratic cone with the unit sphere in $\mathbb{R}^2$. We are interested in the metric properties of this intersection with respect to the intrinsic metric of the unit sphere.

Similarly, a *hyperbolic conic* is the intersection of a quadratic cone with the upper half of the one-sheeted hyperboloid $\{x_0^2 - x_1^2 - x_2^2 = 1\} \subset \mathbb{R}^3$. The Minkowski scalar product induces on the hyperboloid a hyperbolic metric. Again, we are interested in the metric properties of the curve cut by the quadratic cone on the hyperboloid. Central projection from the origin maps the hyperboloid to the disk $x_1^2 + x_2^2 \leq 1$ in the plane $x_0 = 1$, known as the Beltrami–Cayley–Klein model of the hyperbolic plane. The same projection sends a quadratic cone to an affine conic in the plane $x_0 = 1$. Thus, in the Beltrami–Cayley–Klein model (which we shall call, for short, the *Beltrami disk*), a hyperbolic conic is simply the intersection of an affine conic with the disk depicting the hyperbolic plane.

On a more formal and coordinate-free level, a spherical or hyperbolic conic is a pair of quadratic forms (Ω, S) in a 3-dimensional real vector space, such that Ω (the absolute) is non-degenerate.

Non-Euclidean conics share many properties with their Euclidean relatives. For example, the set of points on the sphere with a constant sum of distances from two given points is a spherical ellipse. The same is true in the hyperbolic plane. A reader interested in the bifocal properties of other hyperbolic conics can take a look at Theorem 6.10 or Figure 37.

In spherical geometry, polarity associates to a point the great circle centered at that point. In hyperbolic geometry, the polarity associates to a point outside the Beltrami disk a hyperbolic line polar to that point with respect to the boundary of the disk. (Points inside the disk can be viewed as poles of lines disjoint from the disk.) The polarity exchanges distances and angles. Because of this, for every metric theorem about spherical or hyperbolic conics there is a dual one. For instance, the theorem about constant sum of distances from two given points becomes a theorem about the envelope of lines cutting triangles of constant area from a given angle. See Figures 14 and 36.

The first systematic study of spherical conics was undertaken by Chasles. In the first part of [5] he proves dozens of metric properties of quadratic cones by elegant synthetic arguments. In the second part he interprets them as statements about spherical conics. The work of Chasles was translated from French into English and supplemented by Graves, see also his article [14]. A coordinate approach to spherical conics is used in [24] and [13]. Among the works on hyperbolic conics, the article of Story [23] can be singled out. It uses with much success pencils of conics and the relation between $\Omega(v,w)$ and $\mathrm{dist}(v,w)$. Story's article also contains canonical forms of equations of hyperbolic conics, for which one can also consult [10] and [11].

The present article is to a great extent based on the works of Chasles and Story. We were unable to reproduce all of their results; this could have doubled the length of the article. The interested reader is referred to the originals.

2 Quadratic forms, conics, and pencils

2.1 The dual conic

Let S be a symmetric bilinear form on a real 3-dimensional vector space V. The *isotropic cone* of the corresponding quadratic form is the set

$$I_S = \{v \in V \mid S(v,v) = 0\}.$$

The form S defines a linear homomorphism

$$H_S\colon V \longrightarrow V^*, \quad \langle H_S(v), w\rangle = S(v,w).$$

The symmetry of S translates as the self-adjointness of H_S with respect to the canonical pairing $\langle\cdot,\cdot\rangle$ between V and V^*:

$$\langle H_S(v), w\rangle = \langle v, H_S(w)\rangle.$$

There are one-to-one correspondences between quadratic forms, symmetric bilinear forms, self-adjoint homomorphisms $V \to V^*$, and (complexified) isotropic cones. A *projective conic* (for brevity, we call it simply conic) is an equivalence class of any of the above objects under scaling. By abuse of notation we will often denote a conic by the same symbol as the associated symmetric bilinear form; do not forget that scaling the form does not change the conic.

The elements of $P(V^*)$ can be interpreted as the lines in $P(V)$. Therefore a conic in $P(V)$ is called a *point conic*, and a conic in $P(V^*)$ is called a *line conic*.

With every non-degenerate point conic one can associate a line conic. This can be done in various ways that turn out to be equivalent:

- invert the homomorphism H_S;
- take the image $H_S(I_S)$;
- take the set of tangents to the conic.

Lemma-Definition 2.1. The three constructions listed above result in the same line conic, called the *dual conic* of S and denoted by S^*.

Proof. The first construction yields the cone

$$\{f \in V^* \mid \langle H_S^{-1}(f), f \rangle = 0\},$$

and the second

$$\{H_S(v) \mid v \in V,\ \langle v, H_S(v) \rangle = 0\},$$

which is clearly the same. Also, for every $v \in I_S$ we have

$$\ker H_S(v) = \{w \in v \mid S(v, w) = 0\},$$

which is the plane tangent to I_S along the line spanned by v. □

Thus for the dual conic S^* we have

$$H_{S^*} = H_S^{-1}, \quad I_{S^*} = H_S(I_S).$$

A degenerate point conic is a pair of lines or a double line, and a degenerate line conic is a pair of points or a double point (which means that the conic consists of all lines through one of these points, taken twice if the points coincide). The duality between non-degenerate point conics and line conics does not extend to degenerate ones. The best one can do is to extend the relation $H_S \circ H_{S^*} = \mathrm{Id}$ projectively, see [21, Section 4.D].

2.2 Euclidean vs non-Euclidean geometry

Let Ω be a non-degenerate conic in $P(V)$, hereafter called the *absolute*. The projective transformations from $\mathrm{PGL}(V)$ that map Ω to itself preserve the *Cayley–Klein distance* between the points, defined as half the logarithm of their cross-ratio with the collinear points on the absolute. They also preserve the *Cayley–Klein angles* between the lines, defined through their cross-ratio with the concurrent tangents to Ω (which can be interpreted as the cross-ratio of the corresponding points in $P(V^*)$ with points on the line conic dual to Ω).

For a sign-definite form Ω we obtain the elliptic geometry of $P(V)$; for an indefinite form we obtain hyperbolic (or hyperbolic-de Sitter) geometry.

Euclidean geometry is the geometry of a degenerate line conic that is positive definite and has rank two. Such a conic consists of lines that pass through one of two complex conjugate points. The line through these points is real and is fixed by the transformations that map the conic to itself. This designates it as the line at infinity. A degenerate line conic allows to define the Cayley–Klein angles, but not the Cayley–Klein distances. The corresponding transformation group consists of all similarity transformations.

2.3 Absolute polarity

A non-degenerate absolute Ω corresponds to a self-adjoint isomorphism $H_\Omega \colon V \to V^*$. Projectivization leads to a well-defined isomorphism between projective planes

$$[H_\Omega] \colon P(V) \longrightarrow P(V^*),$$

which, together with its inverse, is called *absolute polarity*. The image of a point $p \in P(V)$ is called the *(absolute) polar* of p and denoted by p°. The preimage of a line $\ell \in P(V^*)$ is called the *(absolute) pole* of ℓ and denoted by ℓ°.

Equivalently, the absolute polar of a linear subspace of V is its orthogonal complement with respect to Ω.

The points $[v], [w] \in P(V)$ such that $\Omega(v, w) = 0$ are called *conjugate* with respect to Ω. In other words, the polar of a point consists of all points conjugate to it.

Two lines are called conjugate if (viewed as elements of $P(V^*)$) they are conjugate with respect to the dual quadratic form Ω^{-1}. Conjugate lines contain the poles of each other.

Absolute polarity does not change when Ω is scaled, thus it is completely defined by the conic Ω. The polar of a point lying on the conic is the tangent at that point. The fact that absolute polarity preserves incidence between points and lines leads to the geometric constructions of poles and polars shown on Figure 1. On this figure, the real part of the isotropic cone of Ω is non-empty, that is, the form Ω is indefinite, which corresponds to hyperbolic geometry. In spherical geometry, the isotropic cone is imaginary. Because of that, a similar geometric construction does not exist, but it can be expressed through an algebraic statement.

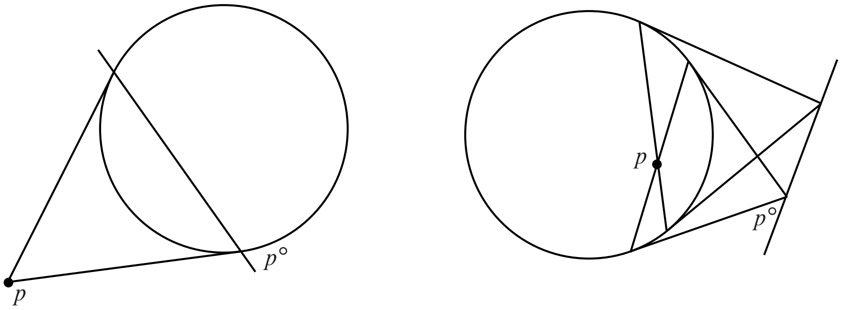


Figure 1. Constructing poles and polars

2.4 The polar conic

Let, in addition to an absolute conic Ω, another non-degenerate conic S be given. Its dual S^*, defined in Section 2.1, is a line conic. We can use absolute polarity to transform it into a point conic. Again, there are three equivalent descriptions of this construction:

- the homomorphism $H_\Omega \circ H_{S^*} \circ H_\Omega \colon V \to V^*$;
- the cone $H_\Omega^{-1}(I_{S^*}) \subset V$;
- the poles of the tangents to the conic S.

Lemma-Definition 2.2. The three constructions listed above result in the same conic, called the *polar conic* of S and denoted by S°.

Proof. The first construction yields the set of all $v \in V$ that satisfy

$$\langle H_\Omega(H_{S^*}(H_\Omega(v))), v\rangle = 0.$$

Since we have

$$\langle H_\Omega(H_{S^*}(H_\Omega(v))), v\rangle = \langle H_{S^*}(H_\Omega(v)), H_\Omega(v)\rangle,$$

this set is the image under H_Ω^{-1} of the set

$$\{f \in V^* \mid \langle H_{S^*}(f), f\rangle = 0\} = I_{S^*}.$$

Hence the second construction is equivalent to the first.

The third construction is equivalent to the second, because the set $H_S(I_S)$ consists of the tangents to the conic S, and the map H_Ω^{-1} represents absolute polarity. ☐

As a consequence, absolute polarity transforms conjugacy with respect to S to conjugacy with respect to S°.

Corollary 2.3. *A line ℓ is the polar of a point p with respect to a conic S if and only if the point ℓ° is the pole of the line p° with respect to the conic S°.*

Proof. The diagram

$$\begin{tikzcd}
P(V^*) \arrow[r, "{[H_{S^*}]}"] \arrow[d, "{[H_\Omega]}"] & P(V) \arrow[d, "{[H_\Omega^{-1}]}"] \\
P(V) \arrow[r, "{[H_{S^\circ}]}"] & P(V^*)
\end{tikzcd}$$

is commutative, since $H_{S^\circ} = H_\Omega \circ H_{S^*} \circ H_\Omega$ by Lemma-Definition 2.2. ☐

2.5 Pencils of conics

Since quadratic forms on a three-dimensional vector space form a vector space of dimension six, projective conics form a five-dimensional projective space.

Definition 2.4. A *pencil of conics* is a line in the projective space of conics.

A pencil spanned by the conics S and T consists of conics of the form $\lambda P + \mu Q$. The degeneracy condition

$$\det(\lambda P + \mu Q) = 0$$

is a homogeneous polynomial of degree 3 in λ and μ. Therefore either all or at most three conics of a pencil are degenerate.

Example 2.5 (conics through four points). Consider a pencil that contains two pairs of lines. If a line of one pair coincides with a line of the other pair, then this line is contained in all conics of the pencil, hence all conics are degenerate. If all four lines are distinct, then the first pair intersects the second pair in four points, and the pencil consists of the conics through these four points, see Figure 2, left. In particular, it contains a third degenerate conic made of a third pair of lines.

Example 2.6 (double contact pencil). Consider a pencil that contains a pair of lines ℓ_1, ℓ_2 and a double line through the points $p_1 \in \ell_1$ and $p_2 \in \ell_2$ (assuming that neither p_1 nor p_2 is the intersection point of ℓ_1 and ℓ_2). Then every conic of the pencil goes through the points p_1 and p_2 and is tangent to the lines ℓ_1 and ℓ_2 at these points (the latter can be seen by viewing the double line as a limit of pairs of lines). See Figure 2, right.

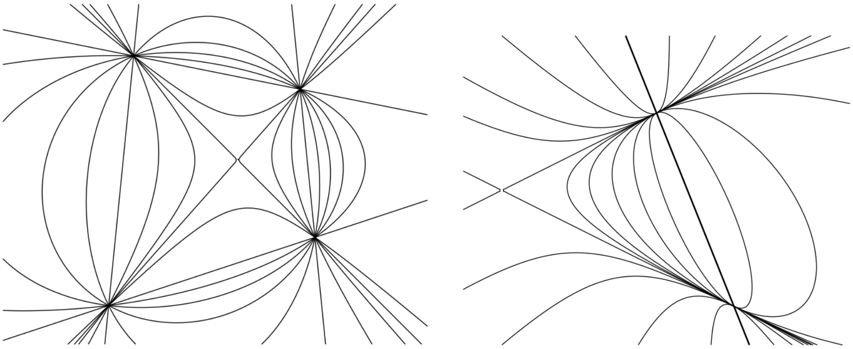


Figure 2. Pencil of conics through four real points and a real double contact pencil

There are pencils that involve imaginary elements, for example the pencil of conics through two pairs of complex conjugate points.

For more details, including other types of pencils with illustrations, see [1, 3.3], [2, 16.4], and [13, 7.3]

2.6 Dual pencils and confocal conics

Take a pencil $\mathcal{P}$ of line conics, that is, a line in the space of conics in $P(V^*)$. The non-degenerate conics of $\mathcal{P}$ can be dualized (see Section 2.1). This gives a family of point conics, called a *dual pencil* or a tangential pencil.

Example 2.7. Let $\mathcal{P}$ be a pencil of line conics through four lines in general position. The corresponding dual pencil consists of non-degenerate point conics tangent to those lines. The degenerate members of $\mathcal{P}$ are three pairs of interection points of the four lines. Thus, in a sense, the dual pencil is spanned by the three diagonals of a complete quadrilateral. Intuitively, these diagonals are degenerate ellipses tangent to the four lines.

Example 2.8. Let $\mathcal{P}$ be a double contact pencil. The dual pencil is also (the non-degenerate part of) a double contact pencil: since the dual conic is made of tangents, the dual of a conic tangent to ℓ at p is a line conic “tangent to” p at ℓ. The degenerate members of this dual pencil are a pair of points (the points of contact) and a double point (the intersection point of the lines of contact).

Definition 2.9. If a pencil of line conics contains the dual absolute, then the corresponding dual pencil is called a *confocal family of conics*.

Recall that in Euclidean geometry there is no absolute point conic, but there is an absolute line conic, see Section 2.2. Therefore Definition 2.9 also makes sense in Euclidean geometry.

We define the foci of spherical and hyperbolic conics in Sections 3.3 and 5.3 and discuss confocal families in Sections 3.5 and 5.7 in more detail.

Lemma 2.10. *Confocal conics intersect orthogonally.*

Proof. This statement means that the tangents of two confocal conics at their intersection point are conjugate with respect to the absolute. Equivalently (in terms of line conics), the points of tangency of a common tangent to two conics of a pencil are conjugate with respect to any conic from this pencil, see Figure 3. But since these points are conjugate with respect to the conics to which they belong, they are so with respect to any linear combination. ☐

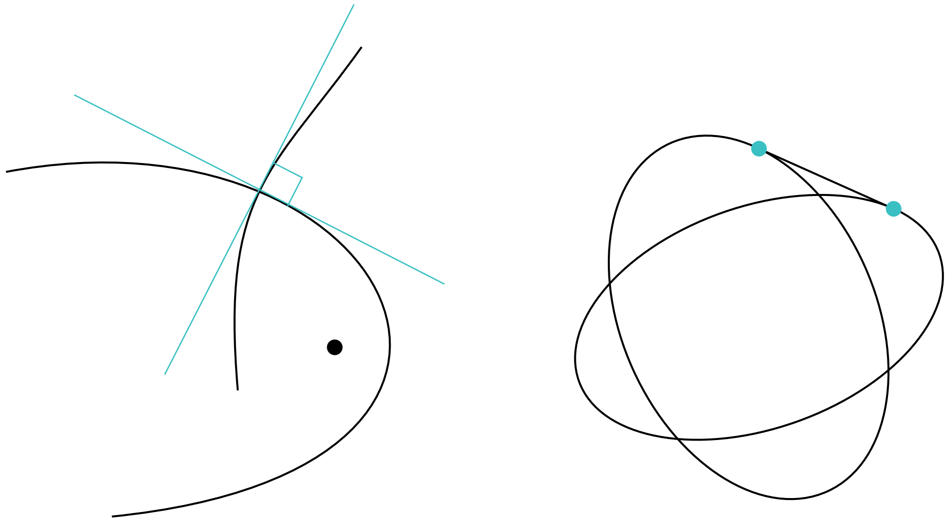


Figure 3. Confocal conics intersect orthogonally, and the dual statement

Example 2.11. A special case of confocal conics is the dual of the pencil spanned by the dual absolute and a double point $2p$. If p does not lie on the absolute, then this is a double contact pencil spanned by the tangents from p to Ω and the double polar of p. This pencil is made by the circles centered at p, which can be shown by a simple computation. (In the hyperbolic case, p can be hyperbolic, de Sitter, or ideal point, see Lemma 6.3.)

2.7 Chasles' theorems

Chasles' article [4] contains a variety of theorems, whose proofs follow all the same principle. A pencil of conics is a line in the projective space of all conics. If in a collection of lines there are many concurrent pairs, then all lines are coplanar, and hence any two of them are concurrent. Translated to pencils of conics, this means that if in a collection of pencils many pairs share a conic, then any two of these pencils share a conic. The argument also works for dual pencils, since they correspond to lines in the space of line conics, and in particular for confocal families of conics.

The following theorem is contained in [4].

Theorem 2.12 (Chasles). *Let* $\{\ell_i\}_{i=1}^4$ *be four tangents to a conic* A. *Denote by* p_{ij} *the intersection point of* ℓ_i *and* ℓ_j. *Assume that the points* p_{12} *and* p_{34} *lie on a conic* B *confocal to* A. *Then the following holds.*

1. *The pairs of points* p_{13}, p_{24} *and* p_{14}, p_{23} *also lie on conics confocal to* A.

2. *The tangents at the points* p_{ij} *to the three conics that contain pairs of points meet at the same point* q.

3. *There is a circle tangent to the lines* ℓ_i, *and this circle is centered at* q.

If, in the hyperbolic case, the point q *is ideal or de Sitter, then the role of a circle centered at* q *is played by a horocycle or a hypercycle.*

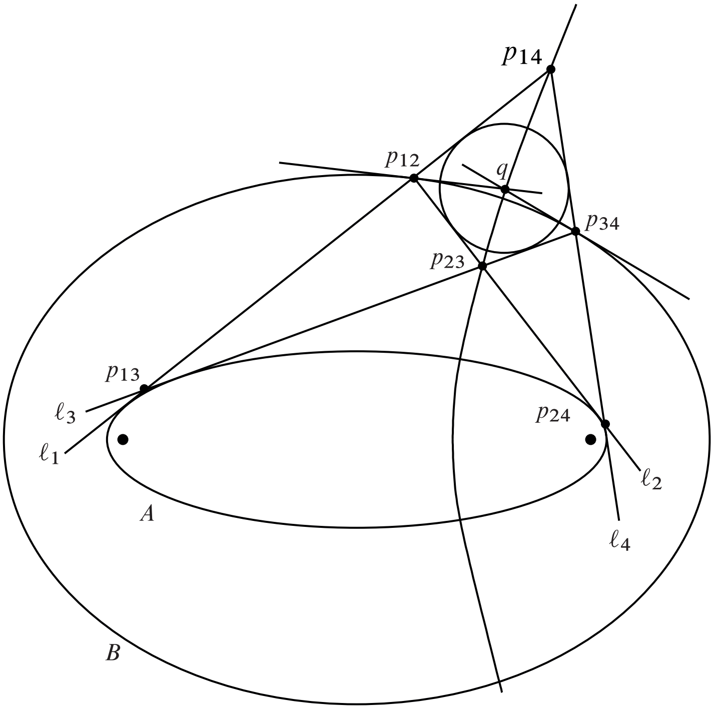


Figure 4. Chasles' theorem

Proof. The idea is to dualize the picture, so that confocal conics become conics collinear with the dual absolute, see Section 2.6.

Consider two pencils of line conics: one spanned by A^* and the dual absolute, the other made by conics through the lines ℓ_i (duals of conics tangent to these lines). The second pencil contains the conics A^*, $p_{12} + p_{34}$, $p_{13} + p_{24}$, and $p_{14} + p_{23}$ (the latter three line conics are degenerate, see the last paragraph of Section 2.1). Since the two pencils share the conic A^*, they lie in a plane in the projective space of line conics.

Consider the pencil of line conics spanned by B^* and $p_{12} + p_{34}$. This is a double contact pencil; it contains the double point $2q$, where q is the intersection point of the tangents to B at p_{12} and p_{34}. See Figure 5, where the line conics are depicted as points.

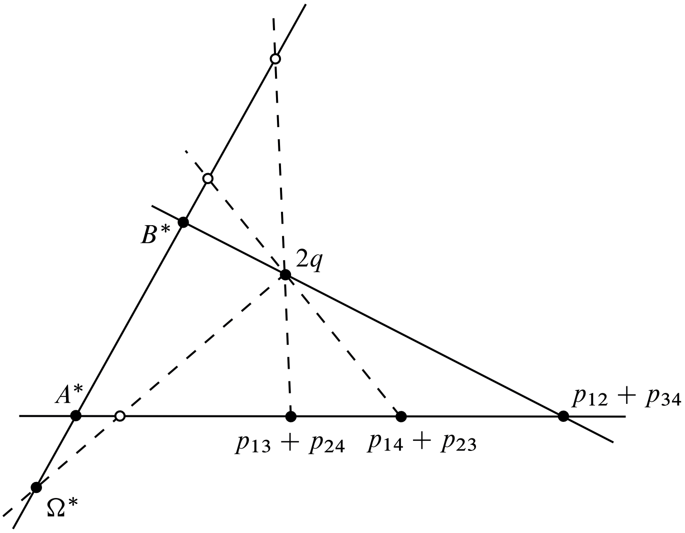


Figure 5. Proof of the Chasles theorem

The pencil of line conics spanned by $p_{14} + p_{23}$ and $2q$ and the pencil spanned by A^* and Ω^* have a conic in common. Its dual is a point conic confocal with A and tangent at the points p_{14} and p_{23} to the lines $p_{14}q$ and $p_{23}q$. The same is true with $p_{13} + p_{24}$ in place of $p_{14} + p_{23}$, which proves the first two parts of the theorem.

For the third part, consider the pencil spanned by Ω^* and $2q$. It consists of circles centered at q (for the hyperbolic case, see Lemma 6.3). This pencil intersects the pencil of conics through the lines ℓ_i, hence there is a circle centered at q tangent to those lines. ☐

The third part of the theorem is related to the following property of confocal conics: a billiard trajectory inside a conic is tangent to a confocal conic.

2.8 Projective properties

The theorems of Pascal, Brianchon, and Poncelet deal with projective properties of conics, therefore they hold for non-Euclidean conics as well as for Euclidean ones. However, in the non-Euclidean case one can obtain new theorems from known ones by modifying them as follows.

- Apply absolute polarity to some of the elements.
- If there is one or several conics in the premises of the theorem, assume one of them to be the absolute conic.
- If there are several conics, assume two of them to be polar to each other.

The theorems below illustrate this.

Theorem 2.13. *The common perpendiculars to the pairs of opposite sides of a spherical or hyperbolic hexagon intersect in a point if and only if the hexagon is inscribed in a conic.*

Proof. The common perpendicular to two lines ℓ_1 and ℓ_2 is the line through the poles ℓ_1° and ℓ_2°. A hexagon is inscribed in a conic if and only if its polar dual is circumscribed about a conic. Hence this theorem follows from the Brianchon theorem by applying absolute polarity. ☐

Theorem 2.14. *The diagonals and the common perpendiculars to the opposite pairs of sides in an ideal hyperbolic quadrilateral meet at a point.*

Proof. The common perpendiculars are the diagonals of the polar quadrilateral, which is circumscribed about the absolute. The diagonals and the lines joining the opposite points of tangency meet at a point; this is a limiting case of the Brianchon theorem. ☐

Theorem 2.15. *The set of points from which a (Euclidean or non-Euclidean) conic is seen under a right angle is also a conic.*

Proof. The tangents drawn from a point to the conic are orthogonal if and only if they harmonically separate the tangents to the absolute conic. This set of points is called the *harmonic locus* of the conic and the absolute. The harmonic locus of any two conics is a conic; this can be proved with the help of Chasles' theory of $(2-2)$ correspondences, see [8, Section 50]. ☐

The set of points from which a curve is seen under a right angle is called the *orthoptic curve*. In the Euclidean case, the orthoptic curve of a conic is a circle.

Theorem 2.16. *Choose any tangent ℓ_1 to a non-Euclidean conic S. Let ℓ_2 be a tangent to S perpendicular to ℓ_1, let ℓ_3 be a tangent to S perpendicular to ℓ_2 and different from ℓ_1, and so on. Assume that for some n we have $\ell_{n+1} = \ell_1$. Then the same holds for any other choice of the tangent ℓ_1.*

First proof. This is the Poncelet theorem for the conic S and its orthoptic conic. ☐

Second proof. The intersection point of the lines ℓ_1 and ℓ_3 is the pole of ℓ_2, which lies on S°. Hence $\ell_1, \ell_3, \ell_5, \dots$ is a Poncelet sequence for the conic S and its polar S°, see Figure 6, left. By assumption, for n odd it closes after n steps, hence it closes after n steps for any choice of ℓ_1. For n even, we have two sequences that close after $\frac{n}{2}$ steps. ☐

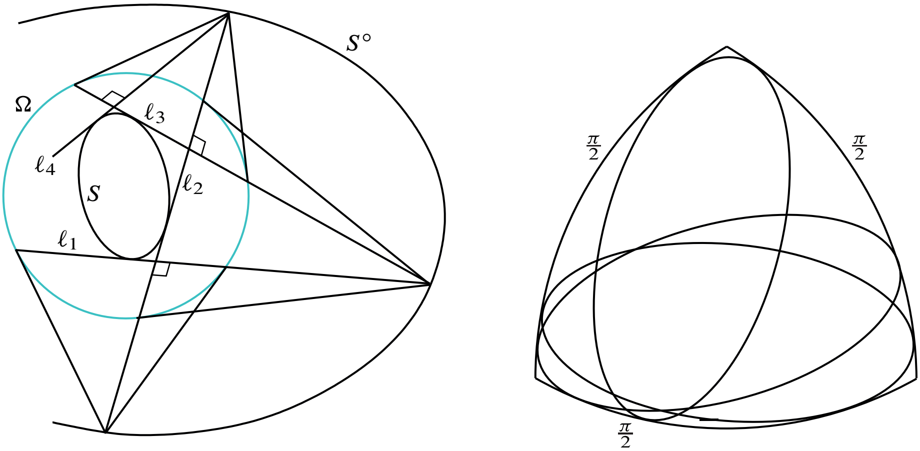


Figure 6. A non-Euclidean variation of the Poncelet theorem

Corollary 2.17. *A spherical ellipse inscribed in a regular right-angled triangle can be rotated while staying inscribed in that triangle. See Figure 6, right.*

Dually, a spherical ellipse circumscribed about a regular right-angled triangle can be rotated while staying circumscribed. See also [13, Theorem 10.1.11].

3 Spherical conics

3.1 A spherical conic and its polar

Algebraically, a spherical conic is a pair of quadratic forms (Ω, S) on a 3-dimensional vector space, with Ω positive definite. In an appropriate basis we have

$$\Omega(v, v) = x^2 + y^2 + z^2.$$

Geometrically, a spherical conic is an intersection of the unit sphere in $\mathbb{R}^3$ with a quadratic cone. If the cone is degenerate, then the conic consists of two great circles or of one "double" great circle. If the cone is circular, then the conic is a pair of diametrically opposite small circles. The most interesting case is when the cone is non-degenerate and non-circular, which we always assume in the sequel.

By the principal axes theorem, there is a basis in which Ω keeps its form, while the quadratic form defining the cone becomes diagonal:

$$C = \{v \in \mathbb{R}^3 \mid S(v, v) = 0\}, \quad S(v, v) = \frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2}, \quad a > b. \tag{1}$$

The polar conic is then given by

$$C^\circ = \{v \in \mathbb{R}^3 \mid S^\circ(v, v) = 0\}, \quad S^\circ(v, v) = a^2x^2 + b^2y^2 - c^2z^2. \tag{2}$$

Following Chasles, we call the z-axis the *principal axis*, the x-axis the *major axis*, and the y-axis the *minor axis* of C. The polar cone has the same principal axis, but the major and the minor axes become interchanged. The points where the axes intersect the sphere are called the *centers* of the conic.

3.2 Projections of a spherical conic

A spherical conic is a spatial curve of degree 4, but from a carefully chosen point of view it takes a more recognizable shape.

Theorem 3.1. *The orthogonal projection of a spherical conic along the principal axis is an ellipse. The projection along the major axis consists of two arcs of a hyperbola, and the projection along the minor axis consists of two arcs of an ellipse. See Figure 7.*

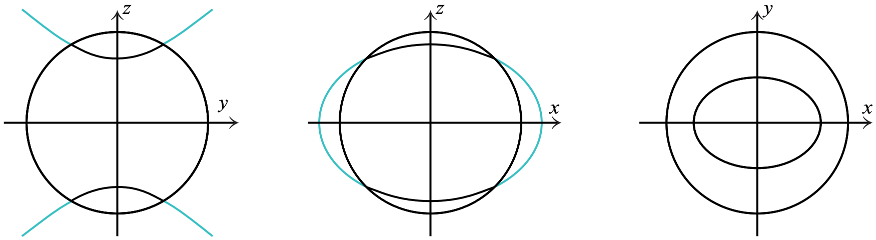


Figure 7. Orthogonal projections of a spherical conic along the axes

Proof. The pencil of affine quadrics spanned by $\Omega = 1$ and $S = 0$ contains the following three cylinders:

$$\left(\frac{1}{a^2} + \frac{1}{c^2}\right)x^2 + \left(\frac{1}{b^2} + \frac{1}{c^2}\right)y^2 = \frac{1}{c^2},$$

$$\left(\frac{1}{a^2} + \frac{1}{c^2}\right)z^2 - \left(\frac{1}{b^2} - \frac{1}{a^2}\right)y^2 = \frac{1}{a^2},$$

$$\left(\frac{1}{b^2} - \frac{1}{a^2}\right)x^2 + \left(\frac{1}{b^2} + \frac{1}{c^2}\right)z^2 = \frac{1}{b^2}.$$

Each of these cylinders intersects the sphere along the conic S. Hence the projection of the conic along the axis of a cylinder is the part of an orthogonal section of the cylinder that lies inside the unit disk. ☐

3.3 Foci and focal lines of a spherical conic

Every quadratic cone has a circular section, a fact that was established by Descartes. Any two parallel sections (not passing through the origin) are similar.

Definition 3.2. A *cyclic plane* of a quadratic cone is a plane through the origin such that every parallel plane intersects the cone in a circle. The intersection of a cyclic plane with the unit sphere is a great circle called a *focal line* of the spherical conic.

A non-circular cone has two cyclic planes, both passing through the major axis and symmetric to each other with respect to the plane spanned by the principal and the major axes.

Definition 3.3. A *cocyclic line* of a quadratic cone is the orthogonal complement of the cyclic plane of the polar cone. The intersection of a cocyclic line with the sphere is an antipodal pair of points called a *focus* of the spherical conic.

A non-circular cone has two cocyclic lines, both of which lie in the plane of the principal and the major axes. Accordingly, a spherical conic has two foci.

Theorem 3.4. *The cyclic planes of the cone* C *from* (1) *are given by the equation*

$$\sqrt{\frac{1}{b^2} - \frac{1}{a^2}}\,y \pm \sqrt{\frac{1}{c^2} + \frac{1}{a^2}}\,z = 0.$$

The cocyclic lines of the cone C *from* (1) *are spanned by the vectors*

$$(\sqrt{a^2 - b^2},\ 0,\ \pm\sqrt{b^2 + c^2}).$$

Proof. The pencil of quadrics $S + \lambda\Omega$ contains a pair of planes

$$S - \frac{1}{a^2}\Omega = \left(\frac{1}{b^2} - \frac{1}{a^2}\right)y^2 - \left(\frac{1}{c^2} + \frac{1}{a^2}\right)z^2.$$

Hence the restriction of S to each of the planes is proportional to the restriction of Ω. This implies that the sections of C parallel to these planes are circular.

The formulas for the cocyclic lines of C follow by computing the cyclic planes of C° and applying polarity (remember that polarity exchanges the major and the minor axes). □

Corollary 3.5. *For every spherical conic, the intersection of its cyclic planes with the plane of the principal and minor axes are the asymptotes of the hyperbola that contains the projection of the conic to that plane. See Figure 7, left.*

Let us now describe some geometric properties of cocyclic lines.

Theorem 3.6. *Two planes through a cocyclic line of a quadratic cone are orthogonal if and only if they are conjugate with respect to the cone.*

Proof. By Corollary 2.3, orthogonality with respect to S translates into orthogonality of the polars with respect to S°. Polars of planes through a cocyclic line of C are lines in a cyclic plane of C°. In this plane, the forms S° and Ω are proportional to each other; in particular, two lines are orthogonal with respect to S if and only if they are orthogonal with respect to Ω. □

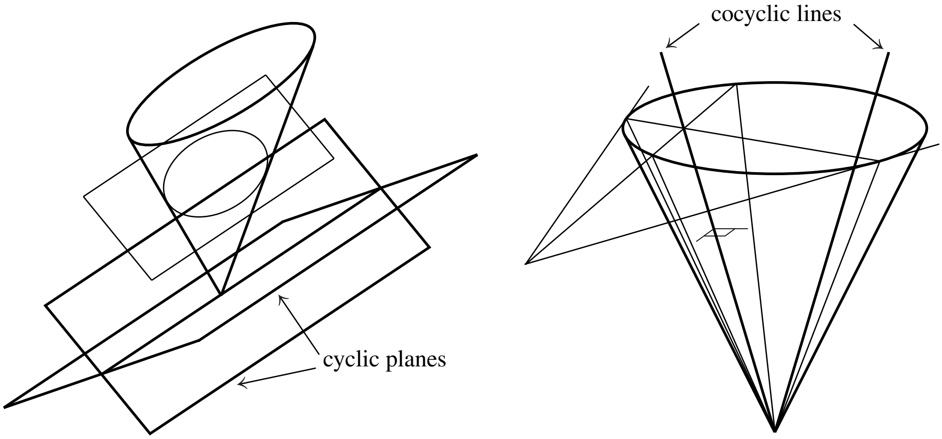


Figure 8. Planes parallel to cyclic planes intersect a cone along a circle. Planes through cocyclic lines are orthogonal if and only if they are conjugate.

Corollary 3.7. *The plane tangent to* $\mathbb{S}^2$ *at a focus* F *of a spherical conic intersects the corresponding cone along an ellipse with a focus at* F.

Proof. Let π be the tangent plane to $\mathbb{S}^2$ at F. Orthogonal planes through a focal line intersect π along orthogonal lines. Thus by the previous theorem two lines in π that pass through F are orthogonal if and only if they are conjugate with respect to the ellipse $C \cap \pi$. This property characterizes the foci of a Euclidean ellipse, see [2, Section 17.2.1.6]. ☐

The following lemma about circular sections will be particularly useful.

Lemma 3.8. *Every sphere passing through a circular section of a quadratic cone intersects it along another circle. Conversely, any two non-parallel circular sections of a quadratic cone are contained in a sphere.*

Proof. Consider the inversion with center at the apex of the cone that sends the sphere to itself. Since the inversion also fixes the cone, it exchanges the two components of the intersection. Therefore, if one component is a circle, so is the other.

For the second part, find an inversion that sends one circular section to the other. Then an invariant sphere passing through one of the circles passes also through the other. ☐

In the limit, as one of the circular sections tends to a point, we obtain the following.

Corollary 3.9. *Every sphere tangent to a cyclic plane at the apex of the cone intersects the cone in a circle.*

3.4 Directors and directrices

Definition 3.10. A line through the origin conjugate with respect to a quadratic cone to one of its cyclic planes is called a *director line* of the cone. The intersection of a director line with the sphere (that is, the pair of poles of a focal line with respect to the conic) is called a *director* of a spherical conic.

In other words, the director lines are formed by the centers of circular sections of the cone.

Definition 3.11. A plane through the origin conjugate with respect to a quadratic cone to one of its cocyclic lines is called a *directrix plane* of the cone. The intersection of a directrix plane with the sphere (that is, the polar of a focus with respect to a conic) is called a *directrix* of a spherical conic.

By Corollary 2.3, the directrix of a conic is polar to the director point of the polar conic.

From the formulas of Theorem 3.4 it follows that the director lines of the cone C are spanned by the vectors

$$\left(0,\, b\sqrt{1-\frac{b^2}{a^2}},\, \pm c\sqrt{1+\frac{c^2}{a^2}}\right),$$

and its directrix planes have the equations

$$\frac{\sqrt{a^2-b^2}}{a^2}x \pm \frac{\sqrt{b^2+c^2}}{c^2}z = 0.$$

3.5 Families of spherical conics

A spherical conic S intersects the absolute conic Ω in four imaginary points that form two complex conjugate pairs (the points are distinct since we assumed the cone to be non-circular).

Theorem 3.12. *The focal lines of a spherical conic are the lines through the complex conjugate pairs of intersection points of the conic with the absolute.*

Proof. This follows from the computation in the proof of Theorem 3.4. But here is an alternative coordinate-free argument. A cyclic plane of a quadratic cone is a plane on which the restrictions of Ω and S are proportional. In particular, Ω and S must vanish at the same time. This implies that a cyclic plane is spanned by two common (imaginary) isotropic lines of Ω and S. Projectively, a cyclic line is a line through two intersection points of Ω and S. For this line to be real, the points must be complex conjugate. ☐

As a consequence, conics that share the focal lines with the conic S form a pencil of conics through four imaginary points, spanned by Ω and S. In coordinates, this pencil is given by the equations

$$\left(\frac{1}{a^2} - \lambda\right)x^2 + \left(\frac{1}{b^2} - \lambda\right)y^2 - \left(\frac{1}{c^2} + \lambda\right)z^2 = 0. \tag{3}$$

One can visualize the conics with common focal lines as follows.

Theorem 3.13. *Every hyperboloid asymptotic to the cone over a spherical conic S intersects the sphere along a spherical conic that shares the focal lines with S.*

Proof. A hyperboloid asymptotic to the cone over S is given by an equation of the form

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = \lambda.$$

By taking a linear combination with the equation of the sphere, we see that the hyperboloid intersects the sphere along the same curve as the cone (3). ☐

By definition of the foci and directrices, we have the following.

Theorem 3.14. *A focus of a spherical conic is the intersection point of two* (*complex conjugate*) *common tangents to the conic and the absolute. The corresponding directrix is the line through the points where these tangents touch the conic.*

That is, conics confocal with S form a dual pencil to the one spanned by S° and Ω and are given by the equations

$$\frac{x^2}{a^2 - \lambda} + \frac{y^2}{b^2 - \lambda} - \frac{z^2}{c^2 + \lambda} = 0, \quad \lambda \in (-1, b^2) \cap (b^2, a^2). \tag{4}$$

Corollary 3.15. *Conics that share a focus and a corresponding directrix form a double contact pencil* (*determined by two complex conjugate points on two complex conjugate lines*).

4 Theorems about spherical conics

4.1 Ivory's lemma

On the turn of the 18th century, Laplace and Ivory [15] computed the gravitational field created by a solid homogeneous ellipsoid. Ivory's solution was immediately and widely recognized for its elegance, see [27], Sections 1141 and 1146. Chasles [4] further simplified Ivory's argument; it is under this form that it is presented nowadays, see e.g. [12], Lecture 30. In particular, Chasles stated explicitely what is now called the Ivory lemma.

Ivory's lemma is a statement about confocal conics and quadrics. Its original version deals with confocal quadrics in Euclidean 3-space, but it holds in any dimension with respect to any Cayley–Klein metric, see [22].

In this section we prove Ivory's lemma on the 2-sphere.

Theorem 4.1 (Ivory's lemma). *The diagonals in a quadrilateral formed by four confocal spherical conics have equal lengths.*

Figure 9 illustrates Ivory's lemma in the Euclidean plane. It can also be interpreted as a distorted view of confocal spherical conics.

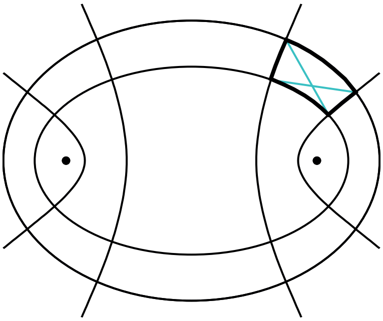


Figure 9. Ivory's lemma

Take our quadratic cone C and a cone C_λ given by equation (4) with $-1 < \lambda < b^2$, which ensures that the corresponding confocal conics do not intersect. Consider a linear map $\mathbb{R}^3 \to \mathbb{R}^3$ given by the diagonal matrix

$$F_\lambda = \operatorname{diag}\left(\frac{\sqrt{a^2-\lambda}}{a}, \frac{\sqrt{b^2-\lambda}}{b}, \frac{\sqrt{c^2+\lambda}}{c}\right).$$

Lemma 4.2. *The map F_λ sends the spherical conic $S = C \cap \mathbb{S}^2$ to the conic $S_\lambda = C_\lambda \cap \mathbb{S}^2$. Besides, if a point $v \in S$ belongs to a conic S_μ, then $F_\lambda(v)$ also belongs to S_μ.*

Proof. Let us set $A = \operatorname{diag}(a^2, b^2, -c^2)$. Then C, C_λ, and C_μ are the isotropic cones of the matrices

$$S = A^{-1}, \quad S_\lambda = (A - \lambda \operatorname{Id})^{-1}, \quad S_\mu = (A - \mu \operatorname{Id})^{-1},$$

respectively. By definition of F_λ we have

$$F_\lambda^2 = (A - \lambda \operatorname{Id}) A^{-1},$$

which immediately implies $F_\lambda(C) = C_\lambda$. Further, observe that

$$\|F_\lambda(v)\|^2 = v^\top (A - \lambda \operatorname{Id}) A^{-1} v = v^\top v - \lambda v^\top A^{-1} v = \|v\|^2 - \lambda v^\top S v = 1,$$

provided that $v \in \mathbb{S}^2 \cap C$. Hence, $v \in S$ implies $v \in S_\lambda$.

We also have

$$\begin{aligned} F_\lambda S_\mu F_\lambda &= A^{-1}(A - \lambda \operatorname{Id})(A - \mu \operatorname{Id})^{-1} \\ &= (A - \mu \Omega)^{-1} - \lambda A^{-1}(A - \mu \operatorname{Id})^{-1} \\ &= (A - \mu \operatorname{Id})^{-1} - \frac{\lambda}{\mu}((A - \mu \operatorname{Id})^{-1} - A^{-1}) \\ &= \frac{\lambda}{\mu} A^{-1} + \Big(1 - \frac{\lambda}{\mu}\Big)(A - \mu \operatorname{Id})^{-1}. \end{aligned}$$

In other words,

$$F_\lambda(v)^\top S_\mu F_\lambda(v) = \frac{\lambda}{\mu} v^\top S v + \Big(1 - \frac{\lambda}{\mu}\Big) v^\top S_\mu v,$$

so that $v \in C \cap C_\mu$ implies $v \in C_\mu$. □

Proof of Theorem 4.1. Let S, S_λ, S_{μ_1}, and S_{μ_2} be confocal conics such that S and S_λ are disjoint, and S_{μ_1}, S_{μ_2} intersect them. Let $v_i \in S \cap S_{\mu_i}$ and $w_i \in S_\lambda \cap S_{\mu_i}$, $i = 1, 2$. We want to show that $\operatorname{dist}(v_1, w_2) = \operatorname{dist}(v_2, w_1)$.

By Lemma 4.2, $w_i = F_\lambda(v_i)$. Since the matrix F_λ is diagonal, we have

$$v_1^\top w_2 = v_1^\top F_\lambda v_2 = w_1^\top v_2,$$

which implies the desired equality of diagonals. □

Remark 4.3. The most general form of the Ivory lemma was discovered by Blaschke (see [3]): it holds in the coordinate nets of Stäckel metrics (in dimension 2 known as Liouville nets). Conversely, if a coordinate net on a Riemannian manifold has the equal diagonals property, then it is a Stäckel net. The argument uses integrals of the geodesic flow and is inspired by Jacobi's ideas from his study of geodesics on ellipsoids. For a modern exposition, see [25] and [16].

4.2 Bifocal properties of spherical conics

On the illustrations to the following theorems we are using the projective model of the sphere: the projection of the sphere from its center to a tangent plane. A spherical conic becomes an affine conic, all great circles (in particular, the focal lines of the conic) become straight lines. As a projection plane, it is convenient to choose a plane that is tangent at one of the centers of the conic, see Figure 10.

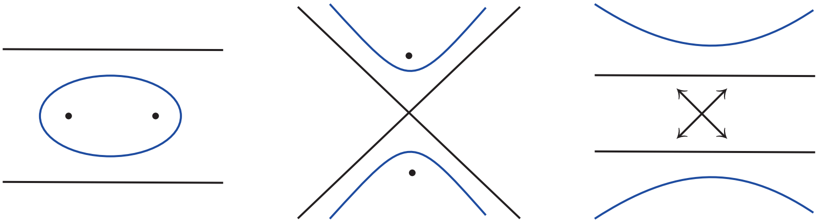


Figure 10. Projections of a spherical conic to the planes tangent to its centers

The focal lines in Figure 10, middle, are not the asymptotes of the hyperbola; the arrows in the center of Figure 10, right, indicate the position of the foci, which belong to the line at infinity.

Each of the theorems below consists of two parts, dual to each other via absolute polarity. Therefore we are proving only one part of each theorem. Usually this is the part that deals with the focal lines, because in the space it deals with circles and spheres and can be proved by an elegant synthetic argument. All proofs were given by Chasles in [5].

Theorem 4.4. 1. *For every tangent to a spherical conic, the point of tangency bisects the segment comprised between the focal lines.*

2. *The lines joining the foci of a spherical conic with a point on the conic form equal angles with the tangent at that point.*

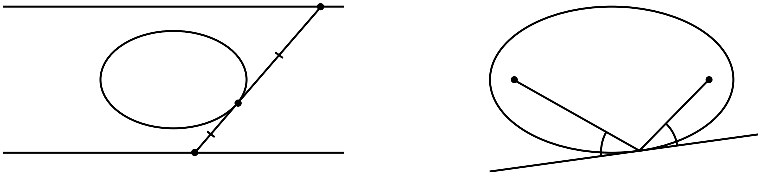


Figure 11. The bisector property

Proof. Interpreted in terms of the corresponding quadratic cone, the first statement becomes as follows: any plane tangent to the cone intersects the cyclic planes along the lines that make equal angles with the line of tangency.

Take two planes parallel to the cyclic planes; they intersect the cone in two non-parallel circles. By Lemma 3.8, these circular sections are contained in a sphere. A tangent plane to the cone intersects the planes of the circles in two lines tangent to the sphere, see Figure 13, left. We need to show that these lines make equal angles with the line of tangency. But any two tangents to the sphere make equal angles with the segment joining their points of tangency, and the theorem is proved. ☐

Theorem 4.4 can be viewed as a limiting case of the following.

Theorem 4.5. 1. *For every secant line of a spherical conic the segments comprised between the conic and the focal lines have equal lengths.*

2. *For every point outside a spherical conic, the angle between a tangent through this point and the line joining the point to a focus is equal to the angle between the other tangent and the line to the other focus.*

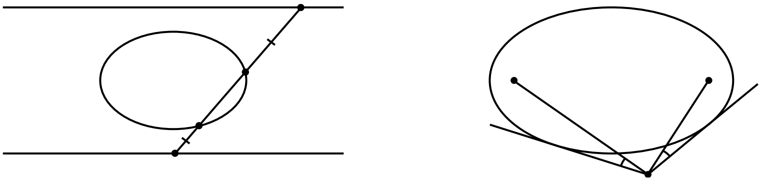


Figure 12. The generalized bisector property

Proof. The first statement says that every plane through two generatrices of the cone intersects the cyclic plane along two lines such that the angle between a line and a generatrix is equal to the angle between the other line and the other generatrix.

Translate the cyclic planes so that they intersect the cone along two circles. By Lemma 3.8 these two circles are contained in a sphere. The generatrices and the intersection lines of the secant plane with the planes spanned by the circles form a planar quadrilateral inscribed in the sphere, see Figure 13, right. Two opposite angles of this quadrilateral complement each other to π, which implies the theorem. ☐

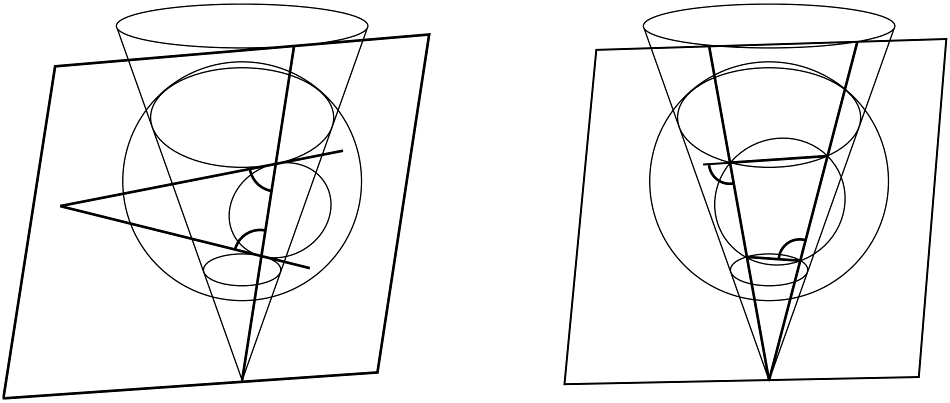


Figure 13. Proofs of Theorems 4.4 and 4.5

Theorem 4.6. 1. *A tangent to a spherical conic cuts from the lune formed by the focal lines a triangle of constant area.*

2. *The sum of the distances from a point on a spherical conic to its foci is constant.*

The first part is dual to the second, because the area of a spherical triangle is equal to its angle sum minus π. Since the angle at the vertex E is constant, the statement is equivalent to the constancy of the sum of the angles at X and Y. These angles are equal to the lengths of F_1Z and F_2Z in the second part of the theorem.

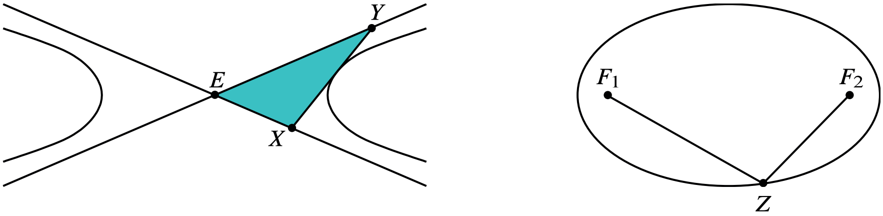


Figure 14. Bifocal properties: Area$(\triangle EXY) =$ const., $F_1Z + F_2Z =$ const.

The second part is due to Magnus [18], who proved it by a direct computation. Two different proofs are given below: the first one uses differentiation, the second one synthetic geometry. The latter is due to Chasles and uses Theorem 4.7 below.

First proof. When we move a line keeping it tangent to the conic, the instantaneous change in the area of the triangle is zero because, by the first part of Theorem 4.4, the point of tangency bisects the segment XY. Similarly, for a point moving along the conic, the derivative of the sum of its distances from the foci is zero by the second part of Theorem 4.4. □

Second proof. The second part of Theorem 4.6 can be derived from the second part of Theorem 4.7 similarly to the proof that in a circumscribed quadrilateral the sums of opposite pairs of sides are equal. Take two points A and B on a conic such that the segments F_1A and F_2B intersect. By Theorem 4.7 there is a circle tangent to the lines F_1A, F_1B, F_2A, F_2B. Since tangent segments drawn from a point to a circle have equal lengths, we have (see Figure 15)

$$F_1A + F_2A = F_1K + F_2L = F_1M + F_2N = F_1B + F_2B. \qquad \square$$

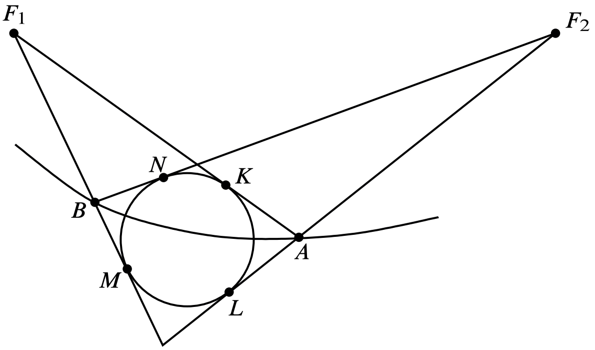


Figure 15. Second proof of Theorem 4.6

Theorem 4.7. 1. *Two tangents to a spherical conic intersect the focal lines in four points that are equidistant from the line through the points of tangency.*

2. *Four lines joining the foci of a spherical conic with two points on the conic are tangent to a circle. The center of this circle is the pole of the line through the two points on the conic.*

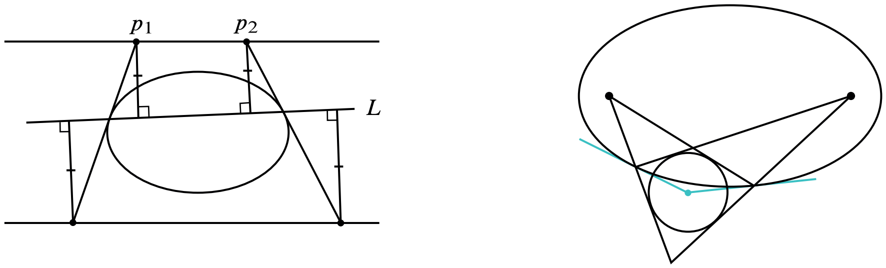


Figure 16. Illustration to Theorem 4.7

Proof. In $\mathbb{R}^3$, the first statement says that for any two tangent planes to the cone their intersection lines with the cyclic planes form equal angles with the plane through the lines of tangency. Theorem 4.4 implies this for the intersection lines of the same tangent plane with different cyclic planes. Let us prove this for the intersection lines of different tangent planes with the same cyclic plane. Figure 17 shows two such lines p_1 and p_2. The shaded triangle lies in the plane L spanned by the tangent lines, compare Figure 16, left.

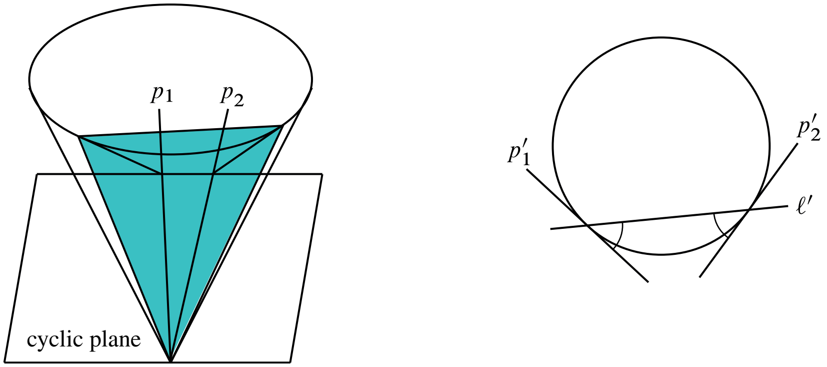


Figure 17. Proof of Theorem 4.7

Translate the cyclic plane parallelly; it will intersect the cone along a circle, the tangent planes along lines p_1' and p_2' parallel to p_1 and p_2, and the plane L along a line ℓ'. The lines p_1' and p_2' make equal angles with the line ℓ', hence they make equal angles with any plane through this line, in particular with L. The theorem is proved. □

The second part of Theorem 4.7 is the third part of Theorem 2.12.

Theorem 4.8. 1. *The product of the sines of distances from the points on a spherical conic to the focal lines is constant.*

2. *The product of the sines of distances from the foci of a spherical conic to its tangents is constant.*

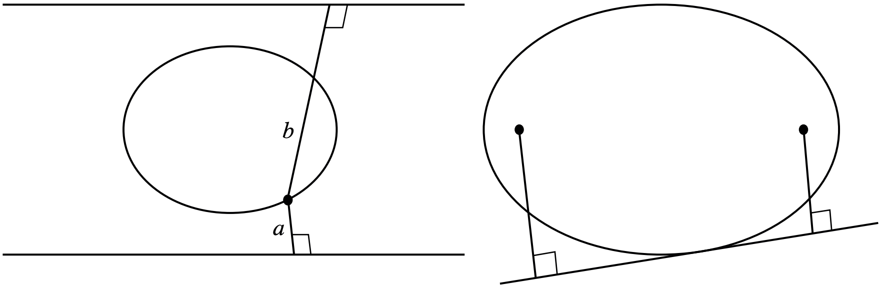


Figure 18. $\sin a \cdot \sin b = \text{const.}$

Proof. Take two non-parallel circular sections of the cone. By Lemma 3.8, there is a sphere through these two sections. Therefore for every generatrix of the cone the product of lengths of the segments between the apex and the circular sections is constant:

$$OX \cdot OY = \text{const.},$$

see Figure 19. On the other hand, we have

$$OX = \frac{OA}{\sin a},$$

$$OY = \frac{OB}{\sin b},$$

where OA and OB are the distances from the apex to the chosen planes, and a, b are the angles between the generatrix and those planes, that is, the distances from the point corresponding to the generatrix to the focal lines of the conic. Since OA and OB are constant, the theorem follows. □

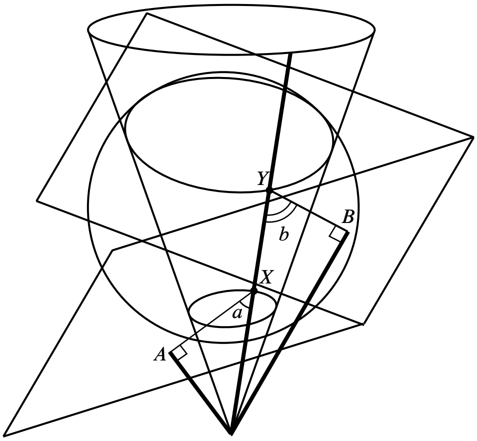


Figure 19. Proof of Theorem 4.8

4.3 The focus-directrix property

Recall that for a point on a Euclidean conic its distance from a directrix is in a constant ratio to its distance from the corresponding focus. The following theorem provides a spherical analog.

Theorem 4.9. 1. *For a tangent to a spherical conic, the sine of its distance from a director point is in a constant ratio to the sine of the angle it makes with the corresponding focal line.*

2. *For a point on a spherical conic, the sine of its distance to a directrix is in a constant ratio to the sine of its distance from the corresponding focus.*

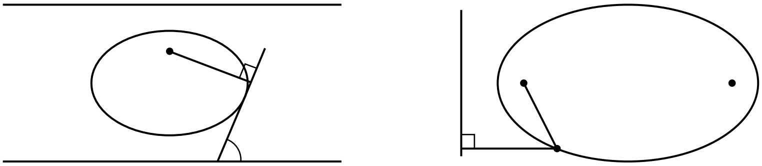


Figure 20. The focus-directrix property and its dual

Proof. Recall that a director point is the intersection point of the sphere with a line formed by the centers of circular sections of a quadratic cone. Take a circular section of the cone. The distance from its center to a plane tangent to the cone is equal to $r \sin a$, where r is the radius of the circle, and a is the angle between the plane of the circle and the tangent plane, see Figure 21. On the other hand, the same distance is equal to $\ell \sin b$, where ℓ is the length of the segment joining the center of the circle to the apex of the cone, and b is the angle between this segment and the tangent plane. Thus we have

$$\frac{\sin a}{\sin b} = \frac{\ell}{r} = \text{const.}$$

At the same time, a is the angle made by the tangent and a focal line, and b is the distance from the corresponding director to that tangent. The theorem is proved. □

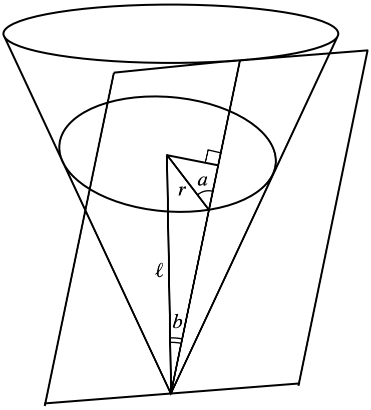


Figure 21. Proof of Theorem 4.9

4.4 Special spherical conics

All spherical conics look essentially the same. However, in certain respects some of them are special.

Theorem 4.10. 1. *The locus of the points from which a spherical arc is seen under a right angle is a spherical conic. The endpoints of the arc belong to the conic and are the poles of its cyclic planes.*

2. *An arc of length* $\frac{\pi}{2}$ *with endpoints moving along two given great circles is tangent to a spherical conic whose foci are the poles of these great circles.*

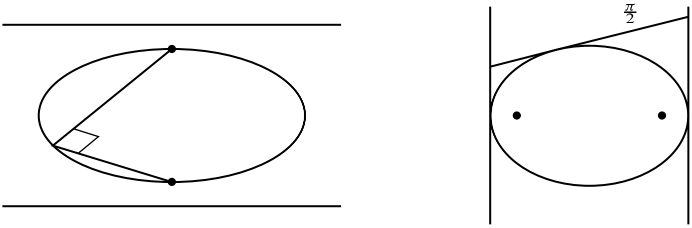


Figure 22. Special spherical conics

Proof. The first statement translates as follows. Choose two lines p_1 and p_2 through the origin and let two planes rotate around these lines while being perpendicular to each other. Then their intersection line describes a quadratic cone whose cyclic planes are orthogonal to p_1 and p_2.

Draw a plane π perpendicular to the line p_1. Planes through p_1 and p_2 are perpendicular if and only if their lines of intersection with π are, see Figure 23. Hence the intersection line of these planes describes a cone over a circle with the segment $[p_1 \cap \pi, p_2 \cap \pi]$ as a diameter. This is a quadratic cone, and the plane π is parallel to one of its cyclic planes. The theorem is proved. ☐

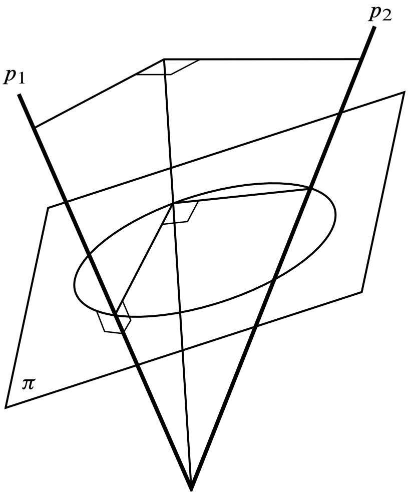


Figure 23. Proof of Theorem 4.9

Lemma 4.11. *The spherical conics from the first part of Theorem 4.10 are given by the equations*

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 0, \quad a > b$$

with $\frac{1}{b^2} = \frac{1}{a^2} + \frac{1}{c^2}$. *The spherical conics from the second part of the same theorem satisfy* $a^2 = b^2 + c^2$.

Proof. By polarity, the statements of the lemma are equivalent. Any one of them can be proved with the help of the formulas from Theorem 3.4. ☐

Another special class of spherical conics is formed by those for which the distance to a focus is equal to the distance to the corresponding directrix. In this respect, they are similar to the Euclidean parabolas.

Theorem 4.12. *The locus of points on the sphere at equal distances from a point and a great circle is a spherical conic with each component of diameter* $\frac{\pi}{2}$. *Equation* (1) *of this conic satisfies* $a = c$.

Proof. That this curve is a component of a spherical conic follows from (the inverse of) the second part of Theorem 4.9 (the focus-directrix property). The two most distant points on the curve are the midpoints of the perpendiculars from the point to the great circle.

For a general conic (1), the diameter of a component is equal to the angle between the rays spanned by the vectors $(a, 0, c)$ and $(-a, 0, c)$. This angle is equal to $\frac{\pi}{2}$ if and only if $a = c$. ☐

Alternatively, one can derive the last theorem from (the inverse of) the second part of Theorem 4.6. For a point F and a great circle d, the condition $\operatorname{dist}(x, F) = \operatorname{dist}(x, d)$ is equivalent to $\operatorname{dist}(x, F) + \operatorname{dist}(x, d^\circ) = \frac{\pi}{2}$, where d° is the pole of d lying in the same hemisphere with respect to d as F. Thus the "spherical parabolas" are also characterized by their foci being the poles of their directrices (the directrix corresponding to a focus must be the pole of the other focus).

For other special spherical conics see [13].

5 Hyperbolic conics

5.1 The hyperbolic-de Sitter plane

Let Ω be a quadratic form on $\mathbb{R}^3$ of signature $(-, +, +)$. The hyperbolic plane is a component of the hyperboloid of two sheets $\Omega(x, x) = -1$, equipped with a

Riemannian metric induced by the form Ω. In the Beltrami–Cayley–Klein model, the hyperboloid is projected from the origin to an affine plane and becomes the interior region of a conic, the *absolute conic*. Points on the absolute are called *ideal* or *absolute points*. Geodesics in the Beltrami–Cayley–Klein model are straight line segments with ideal endpoints. The geodesic distance between two points is half the logarithm of their cross-ratio with the collinear ideal points.

It is convenient to view the plane of the Beltrami–Cayley–Klein model as a projective plane; it is then nothing else but the projectivization of $\mathbb{R}^3$. On pictures, we show an affine chart of this plane, with the absolute in the form of a circle.

The exterior of the absolute, homeomorphic to an open Möbius band, is called the *de Sitter plane*. Polarity with respect to the absolute conic sends hyperbolic points to de Sitter lines (projective lines disjoint from the absolute), ideal points to lines tangent to the absolute, and de Sitter points to hyperbolic lines.

For more details on the de Sitter geometry, see [9].

5.2 Classification of hyperbolic conics

Algebraically, a hyperbolic conic is a pair of quadratic forms (Ω, Q) in $\mathbb{R}^3$, where the absolute form Ω has signature $(-,+,+)$. We assume Q to be indefinite (thus with non-empty isotropic cone) and usually non-degenerate, so that without loss of generality it has signature $(-,+,+)$ as well.

Geometrically, in the Beltrami–Cayley–Klein model, a hyperbolic conic is the part of an affine conic inside the absolute. The part outside the absolute may be called a de Sitter conic. However, it is convenient to consider both parts at the same time. Under absolute polarity, the hyperbolic (respectively, de Sitter) points of a conic correspond to the tangents at the de Sitter (respectively, hyperbolic) points of the polar conic.

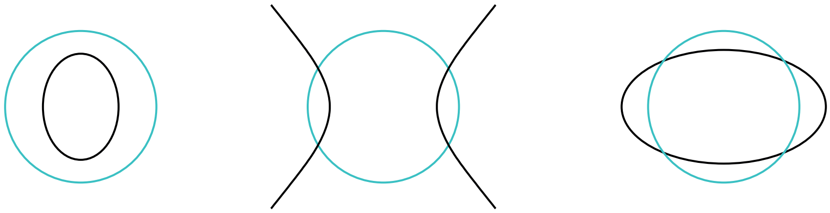
For a pair of indefinite quadratic forms the principal axes theorem in general does not hold. A geometric manifestation of this is the variety of different relative positions of two real conics, and hence the variety of different types of hyperbolic conics. Following Klein [17], we classify hyperbolic conics according to the multiplicity and the reality of their ideal points.

Definition 5.1. An intersection point of a hyperbolic conic with the absolute $\Omega = 0$ is called an *absolute point* of the conic. A common tangent to the conic and the absolute is called an *absolute tangent* to the conic.

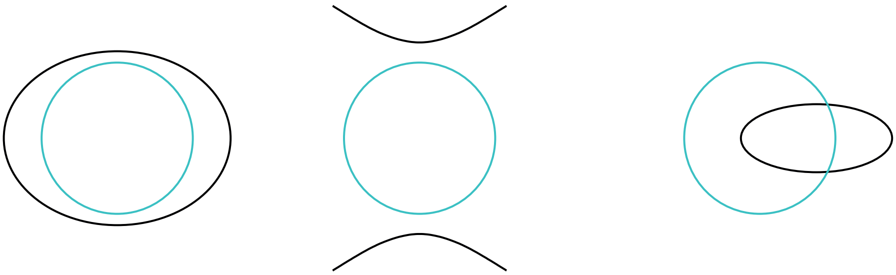
There are four absolute points and four absolute tangents, counted with multiplicity and including imaginary elements. Imaginary points or lines come in conjugate pairs.

Non-degenerate hyperbolic conics are subdivided into ellipses, hyperbolas, parabolas, and cycles. Ellipses and hyperbolas (see Figure 24) have four distinct absolute points. Parabolas (see Figure 25) have at least one simple and at least one multiple absolute point. Finally, cycles (see Figure 26) have either two double ore one quadruple absolute point.

Conics occupying the same column on Figures 24–26 are dual to each other, with the exception of concave hyperbolas and semihyperbolas, which are self-dual. Another self-dual class of conics are osculating parabolas.



Ellipse Convex hyperbola Concave hyperbola



De Sitter Ellipse De Sitter hyperbola Semihyperbola

Figure 24. Hyperbolic ellipses and hyperbolas

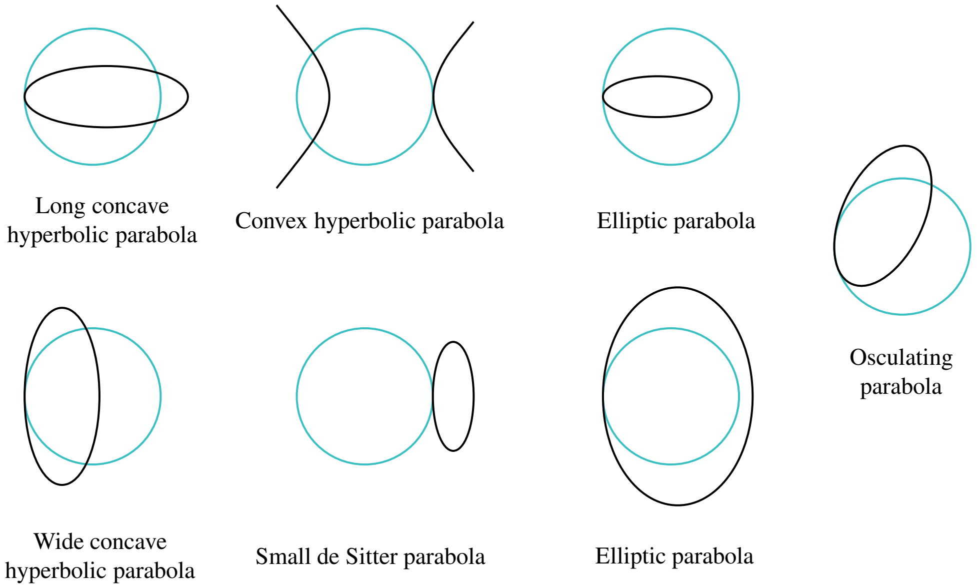


Figure 25. Hyperbolic parabolas

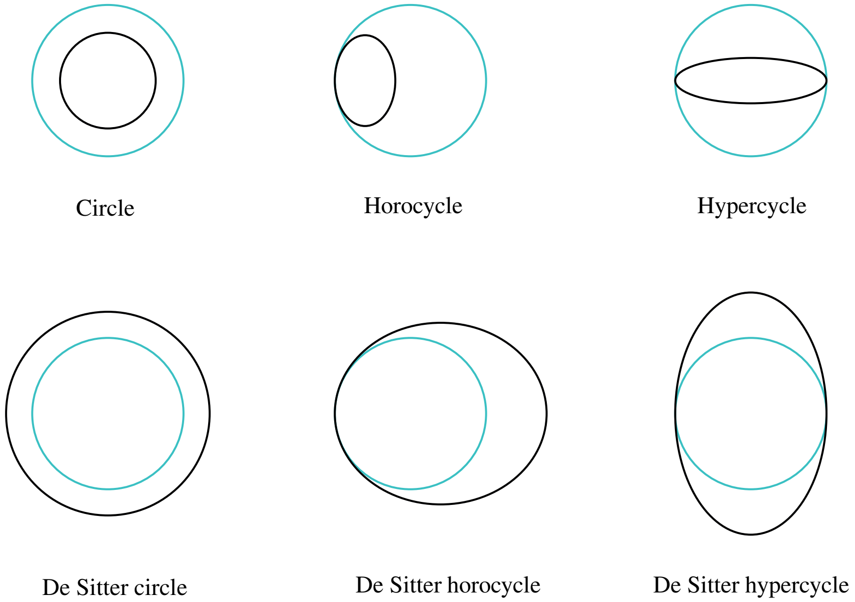


Figure 26. Hyperbolic circles

5.3 Foci and focal lines of a hyperbolic conic

Definition 5.2. A *focal line* of a conic is a line through two of its absolute points. A *focus* of a conic is an intersection point of two of its absolute tangents.

If a conic is tangent to the absolute, then the point of tangency is considered as a focus, and the tangent at this point as a focal line of the conic.

A focal line and a focus are dual notions: the pole of a focal line is a focus of the polar conic.

Two complex conjugate lines intersect in a real point, and two complex conjugate points lie on a real line. Therefore every conic has at least one pair of (possibly coincident) real foci and at least one pair of (possibly coincident) real focal lines.

Lemma 5.3. *A real focus of a hyperbolic conic is a hyperbolic, ideal, or a de Sitter point at the same time as it lies inside, on the boundary, or outside of the oval bounded by the conic.*

Proof. This is quite obvious, and becomes even more obvious in the dual formulation: a real focal line is hyperbolic, ideal, or de Sitter at the same time as it intersects, is tangent, or is disjoint from the conic. ☐

Ellipses and hyperbolas have three pairs of focal lines and three pairs of foci, every pair corresponding to different matchings of four absolute points or four absolute lines. All foci of concave hyperbolas and de Sitter hyperbolas are real and lie in the de Sitter plane. The other ellipses and hyperbolas have only one pair of real foci.

Non-osculating parabolas have two pairs of foci. One pair is real and contains an ideal point. The other pair is double, that is, it splits in two pairs if the parabola is perturbed so that to become an ellipse or a hyperbola. The double pair of foci is real for the concave parabolas and for the de Sitter parabola. The osculating parabola has a triple pair of foci, consisting of the point of tangency and of a de Sitter point on the corresponding absolute tangent.

All cycles have a pair of coincident foci: the center of a circle, the ideal point of a horocycle, and the pole of the center line of a hypercycle. Hypercycles have in addition a double pair of foci: the endpoints of the center line.

Lemma 5.4. *A conic and the absolute induce the same Cayley–Klein metric on each focal line and on the pencil of lines through each focus.*

Proof. A Cayley–Klein distance between two points is determined by their cross-ratio with the two (possibly imaginary) collinear points on the conic. A focal line intersects the conic and the absolute in the same points, therefore the two metrics on this line coincide.

Similarly, a Cayley–Klein angle between two lines is determined by the cross-ratio of these lines with the two tangents from the same pencil. In the line pencil through a focus the same two lines are tangent to the conic and to the absolute. ☐

Compare this with Theorems 3.12 and 3.14 in the spherical case.

Lemma 5.4 implies the following property of a focus of a hyperbolic conic: for any line through the focus, its absolute pole and the pole with respect to the conic are collinear with the focus, see Figure 27, left. The Euclidean analog of this is shown on Figure 27, right.

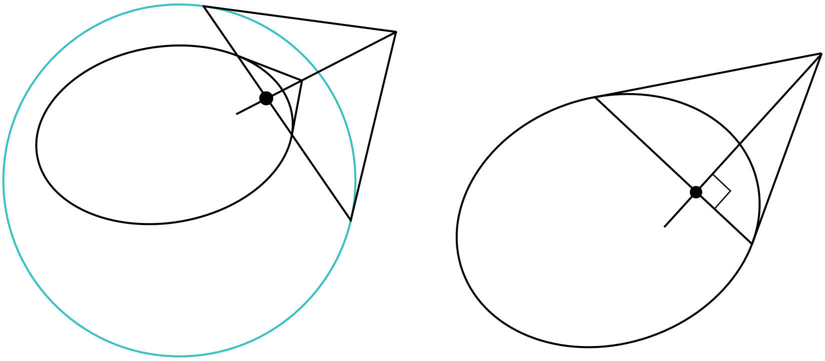


Figure 27. Characteristic properties of foci of hyperbolic and Euclidean conics

5.4 Axes and centers

As we already noted, foci and focal lines come in pairs.

Definition 5.5. The intersection of a pair of focal lines is called a *center* of the conic. The line through a pair of foci is called an *hyperbolic conic!axis* of the conic.

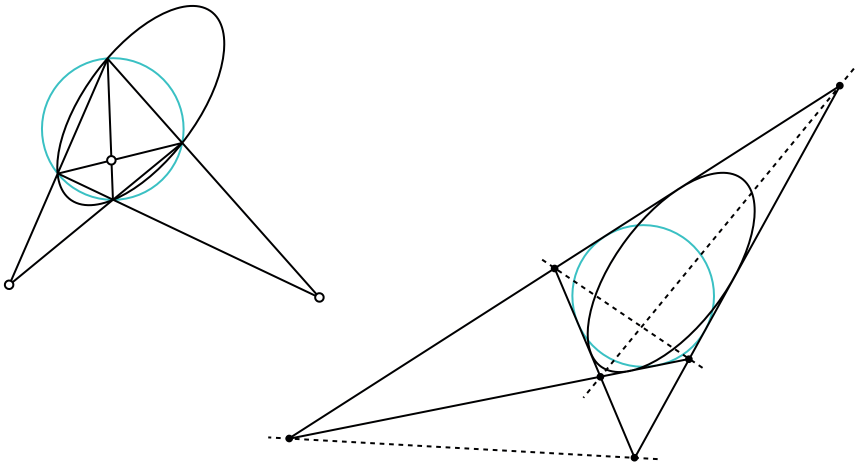


Figure 28. Centers and axes of a concave hyperbola

All hyperbolic ellipses and hyperbolas, except the semi-hyperbola, have three distinct centers and three distinct axes. The case of a concave hyperbola is illustrated on Figure 28.

Lemma 5.6. 1. *The centers of a conic are pairwise conjugate with respect to the conic as well as with respect to the absolute.*

2. *The axes of a conic are pairwise conjugate with respect to the conic as well as with respect to the absolute.*

3. *A line through two centers is an axis; an intersection point of two axes is a center.*

Proof. It is a classical fact that the three diagonal points of a quadrangle inscribed in a conic are pairwise conjugate with respect to the conic, see e.g. [2, Section 14.5.2]. The quadrangle of the absolute points is inscribed both in the conic and in the absolute. This implies the first part of the lemma.

The second part is dual to the first one.

Three pairwise conjugate lines or three pairwise conjugate points form a self-polar triangle. Since two conics in general position have a unique common self-polar triangle, it follows that the centers and the axes span the same triangle. ☐

The third part of the lemma has the following reformulation: for two conics meeting in four points, the diagonals of the common inscribed quadrilateral and those of the common circumscribed one meet at the same point. An alternative proof of this is to apply a projective transformation that sends the four intersection points to the vertices of a square.

5.5 Directors and directrices

Definition 5.7. The pole of a focal line with respect to the conic is called a *director point* of the conic. The polar of a focus with respect to the conic is called a *directrix* of the conic.

Compare Definitions 3.10 and 3.11.

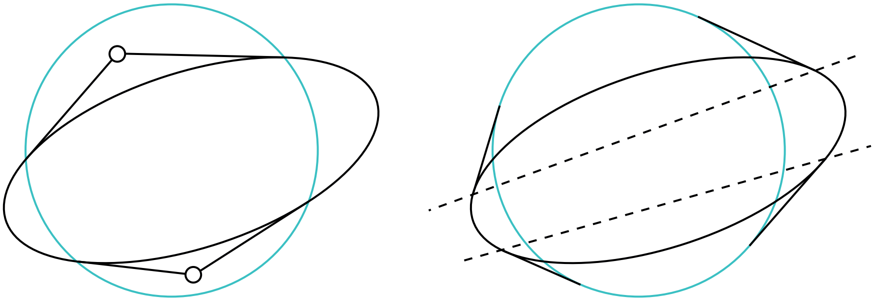


Figure 29. A pair of director points and a pair of directrices

Lemma 5.8. *Director points of a conic lie on its axes. Directrices of a conic pass through its centers.*

Proof. Draw a line through a pair of director points. By the basic properties of polarity, the pole of this line with respect to the conic is the intersection point of the corresponding focal lines, see Figure 30, that is, a center of the conic. It follows that the line through a pair of director points is an axis.

The second part of the lemma is dual to the first. ☐

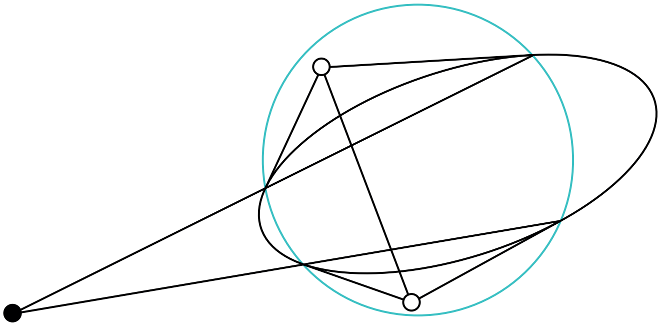


Figure 30. A pair of director points span an axis

Corollary 5.9. *Through every center of a hyperbolic conic pass two focal lines, two axes, and two directrices.*

5.6 Examples

Let us locate foci and directrices for some types of hyperbolic conics. We assume that $\Omega(v, v) = x^2 + y^2 - z^2$, so that in the Beltrami–Cayley–Klein model in the plane $z = 1$ the absolute is the unit circle centered at the origin.

Example 5.10. By a linear transformation that preserves the absolute, a hyperbolic ellipse can be brought to a canonical form

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \quad 1 > a > b.$$

The real foci, real focal lines, and real directrices are

$$\Big(\pm\sqrt{\frac{a^2 - b^2}{1 - b^2}}, 0\Big), \quad x = \pm a\sqrt{\frac{1 - b^2}{a^2 - b^2}}, \quad x = \pm a^2\sqrt{\frac{1 - b^2}{a^2 - b^2}}.$$

The directrices are tangent to the absolute if and only if $b = \frac{a}{\sqrt{1+a^2}}$. For smaller b the directrices are hyperbolic lines, for larger b they are de Sitter lines, see Figure 31.

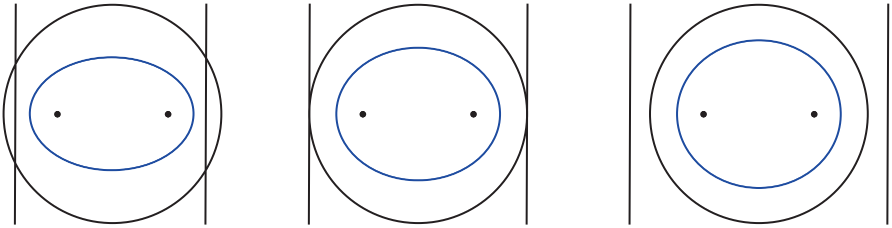


Figure 31. Foci and directrices of hyperbolic ellipses

Example 5.11. A semi-hyperbola can be brought to a canonical form

$$ax^2 - 2x + \frac{y^2}{b^2} = 0, \quad |a| < 2.$$

The only pair of real foci is

$$F_1 = \Big(\frac{b^2}{1 - \sqrt{b^4 - ab^2 + 1}}, 0\Big), \quad F_2 = \Big(\frac{b^2}{1 + \sqrt{b^4 - ab^2 + 1}}, 0\Big).$$

The focus F_1 is a de Sitter point, the focus F_2 is a hyperbolic point. The corresponding directrices d_1 and d_2 are given by equations

$$x = \frac{b^2}{ab^2 - 1 + \sqrt{b^4 - ab^2 + 1}}, \quad x = \frac{b^2}{ab^2 - 1 - \sqrt{b^4 - ab^2 + 1}},$$

respectively. For $a = \frac{1}{b^2}$ (that is, when the semi-hyperbola is represented by a circular arc) both directrices are tangent to the absolute, for $a < \frac{1}{b^2}$ the directrix d_1 is de Sitter while d_2 is hyperbolic, finally for $a > \frac{1}{b^2}$ the directrix d_1 is hyperbolic while d_2 is de Sitter, see Figure 32.

If $a = 0$ (the semi-hyperbola is represented by a parabola), then $d_1 = F_2^\circ$, and $d_2 = F_2^\circ$. In this case the semi-hyperbola is the locus of points equidistant from the point F_1 and the line d_1, see Example 6.6.

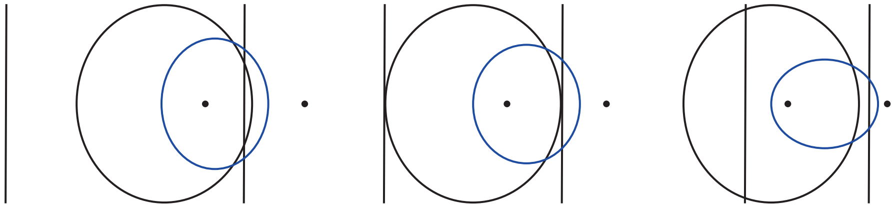


Figure 32. Foci and directrices of semihyperbolas

Example 5.12. An elliptic parabola can be brought to a canonical form

$$\frac{(x-(1-a))^2}{a^2} + \frac{y^2}{b^2} = 1, \quad b^2 < a < 1.$$

(The condition $b^2 < a$ ensures that the curve stays inside the unit circle.) The only pair of real foci:

$$F_1 = \left(1 - \frac{2(a-b^2)}{1-b^2}, 0\right), \quad F_2 = (1, 0).$$

The corresponding directrices are given by the equations

$$d_1 = \left\{x = 1 + \frac{2a(a-b^2)}{-a+2b^2-ab^2}\right\}, \quad d_2 = \{x = 1\}.$$

The directrix d_1 is tangent to the absolute for $a = 2b^2$, hyperbolic for $a > 2b^2$, and de Sitter for $a < 2b^2$.

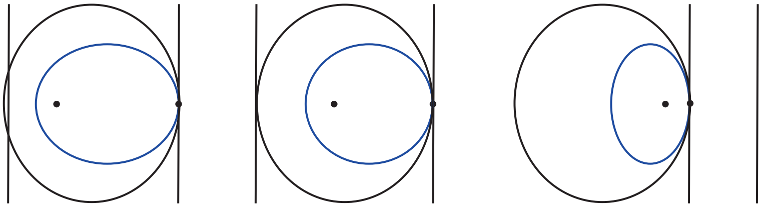


Figure 33. Foci and directrices of elliptic parabolas

5.7 Families of hyperbolic conics

Since the focal lines of a hyperbolic conic are determined by its intersection points with the absolute, the conics that share the focal lines form a pencil of conics containing the absolute conic. For ellipses and hyperbolas, this pencil is determined by four distinct points (some of which can form complex conjugate pairs). For example, if all four points are real, then the pencil is formed by convex and concave hyperbolas and by three pairs of focal lines.

By polarity between foci and focal lines, the confocal conics (those sharing a pair of foci) form a dual pencil. There are several types of confocal nets of hyperbolic conics. The simplest ones are formed by concentric cycles and pencils of lines through their centers. Confocal nets formed by ellipses, hyperbolas, or parabolas are shown on Figures 34 and 35.

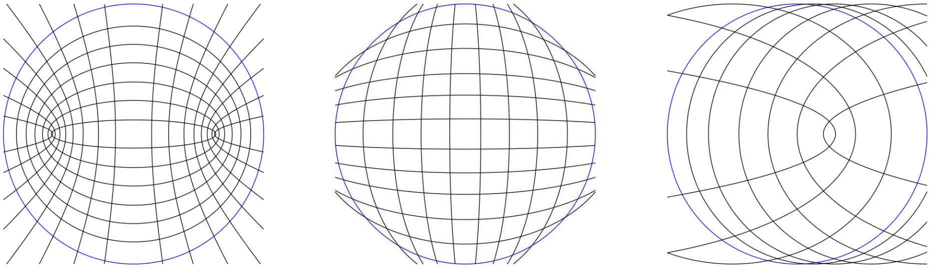


Figure 34. Confocal hyperbolic ellipses and hyperbolas

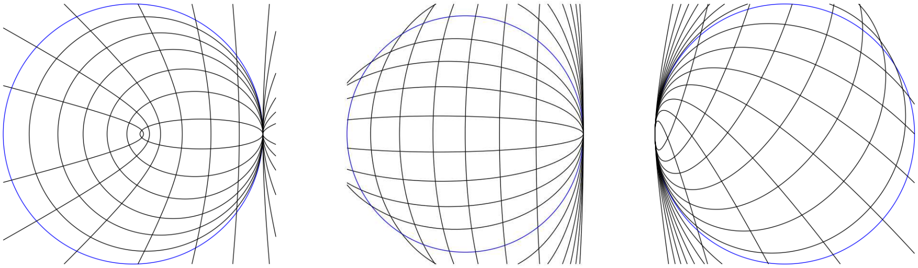


Figure 35. Confocal hyperbolic parabolas

Conics that share a focus and a corresponding directrix form a double contact pencil, compare Corollary 3.15 in the spherical case.

6 Theorems about hyperbolic conics

6.1 Ivory's lemma

Similarly to the Euclidean and to the spherical case (see Theorem 4.1), we have the following.

Theorem 6.1 (Ivory's lemma). *The diagonals in a quadrilateral formed by four confocal hyperbolic conics have equal lengths.*

This theorem is proved in [22] by a method similar to that used in the spherical case, see Section 4.1: one finds a homomorphism $F_\lambda\colon \mathbb{R}^3 \to \mathbb{R}^3$ that is self-adjoint with respect to Ω and maps one side of the quadrilateral onto the opposite side.

6.2 Geometric interpretations of the scalar product

This section deals with metric properties of hyperbolic conics. In the proofs it will be more convenient to use the hyperboloid model of the hyperbolic plane. We have

$$\{x \in \mathbb{R}^3 \mid \Omega(x,x) = -1\} = \mathbb{H}^2 \cup (-\mathbb{H}^2),$$

where $\mathbb{H}^2$ and $-\mathbb{H}^2$ are two components of the hyperboloid. The distance between two points $x, y \in \mathbb{H}^2$ can be computed by the formula

$$\Omega(x,y) = -\cosh(\operatorname{dist}(x,y)).$$

The hyperboloid of one sheet represents a double cover of the de Sitter plane:

$$\{x \in \mathbb{R}^3 \mid \Omega(x,x) = 1\} = \widetilde{d\mathbb{S}^2}.$$

As before, a de Sitter point $x \in \widetilde{d\mathbb{S}^2}$ is the pole of a hyperbolic line

$$x^\circ = \{y \in \mathbb{H}^2 \mid \Omega(x,y) = 0\}.$$

An advantage of $\widetilde{d\mathbb{S}^2}$ over $d\mathbb{S}^2$ is that it allows to give a simple description of hyperbolic half-planes:

$$x^\circ_+ = \{y \in \mathbb{H}^2 \mid \Omega(x,y) \geq 0\}, \quad x^\circ_- = \{y \in \mathbb{H}^2 \mid \Omega(x,y) \leq 0\}.$$

The scalar product of two de Sitter points computes the angle or the distance between the corresponding hyperbolic lines, and the scalar product of a hyperbolic and a de Sitter point computes the distance between a point and a line. For details, see Lemma 6.2 below.

Finally, there is the isotropic cone, which we divide in two parts

$$\{x \in \mathbb{R}^3 \mid \Omega(x,x) = 0\} = L \cup (-L),$$

where L and $-L$ are one-sided cones asymptotic to $\mathbb{H}^2$ and $-\mathbb{H}^2$, respectively. For $x \in L$, its polar x° is the plane through x tangent to the isotropic cone, which corresponds to a line tangent to the absolute. But there is a more interesting object that one can associate with a point of L. Define

$$H_x = \{y \in \mathbb{H}^2 \mid \Omega(x,y) = -1\}.$$

Then H_x is a horocycle centered at x (it is not hard to show that the normals to H_x pass through x). As a curve in $\mathbb{R}^3$, H_x is a parabola; the central projection makes it an ellipse in the Beltrami–Cayley–Klein model. This ellipse osculates the absolute at the image of the point x. The scalar product $\Omega(x,y)$ for $x \in \mathbb{H}^2$ and $y \in L$ measures the distance between x and H_y.

To summarize, every vector in $\mathbb{H}^2 \cup L \cup \widetilde{d\mathbb{S}^2} \subset \mathbb{R}^3$ corresponds to a geometric object in the hyperbolic plane: a point, a (co-oriented) line, or a horocycle; the scalar product of two vectors measures the distance between the corresponding objects. An exact formulation is given in the lemma below.

Lemma 6.2. 1. *If $x, y \in \mathbb{H}^2$, then*

$$\Omega(x,y) = -\cosh \operatorname{dist}(x,y).$$

2. *If $x \in \mathbb{H}^2$ and $y \in \widetilde{d\mathbb{S}^2}$, then*

$$\Omega(x,y) = \sinh \operatorname{dist}(x, y^\circ),$$

where the point-to-line distance $\operatorname{dist}(x, y^\circ)$ *is taken positive if the vectors x and y lie on the same side from the plane $y^\circ \subset \mathbb{R}^3$, and negative otherwise.*

3. *If $x \in \mathbb{H}^2$ and $y \in L$, then*

$$\Omega(x,y) = -e^{\operatorname{dist}(x,H_y)},$$

where the distance to a horocycle is taken to be negative for the points inside the horocycle and positive for the points outside.

4. *If* $x, y \in \widetilde{d\mathbb{S}^2}$, *then*

$$\Omega(x, y) = \begin{cases} \cos(\angle(x^\circ, y^\circ)) & \text{if } x^\circ \cap y^\circ \text{ is hyperbolic,} \\ \pm 1 & \text{if } x^\circ \cap y^\circ \text{ is ideal,} \\ \pm \cosh(\operatorname{dist}(x^\circ, y^\circ)) & \text{if } x^\circ \cap y^\circ \text{ is de Sitter.} \end{cases}$$

Here $\angle(x^\circ, y^\circ)$ *is an angle between the lines* x° *and* y° *occupied by the points* z *where* $\Omega(x, z)$ *and* $\Omega(y, z)$ *have different signs.*

5. *If* $x \in \widetilde{d\mathbb{S}^2}$ *and* $y \in L$, *then*

$$\Omega(x, y) = e^{\operatorname{dist}(x^\circ, H_y)},$$

where the distance between a line and a horocycle is the length of the common perpendicular taken with the minus sign if the line intersects the horocycle.

6. *If* $x, y \in L$, *then*

$$\Omega(x, y) = -2e^{\operatorname{dist}(H_x, H_y)},$$

where the distance between two horocycles is the length of the common perpendicular taken with the minus sign if the horocycles intersect.

This is proved via an appropriate parametrization of the line spanned by the points x and y, see e.g. [20] and [26].

As a simple application of the previous lemma, we can describe the quadratic cones that correspond to hyperbolic cycles.

Lemma 6.3. *For every* $c \in \mathbb{R}$ *and* $p \in \mathbb{R}^3$, *the quadratic cone*

$$\Omega(x, x) + c \cdot \Omega(x, p)^2 = 0$$

corresponds in the hyperbolic-de Sitter plane with the absolute Ω *to a cycle centered at the point* p. *Namely, it describes*

$$\begin{cases} \text{a circle with the center } p & \text{if } \Omega(p, p) < 0, \\ \text{a horocycle with the ideal point } p & \text{if } \Omega(p, p) = 0, \\ \text{a hypercycle around the line } p^\circ & \text{if } \Omega(p, p) > 0. \end{cases}$$

Proof. The intersection of this quadratic cone with $\mathbb{H}^2 \subset \mathbb{R}^3$ is formed by the points x with $\Omega(x, p) = \text{const}$. Lemma 6.2 implies that these are the points at a constant distance from p, if $p \in \mathbb{H}^3$, or the points at a constant distance from a horocycle centered at p (hence also a horocycle), if $p \in L$, and the points at a constant distance from the line p°, if $p \in \widetilde{d\mathbb{S}^2}$. ☐

Results of the following sections are essentially due to Story [23]. However, he does not care about the position of foci, focal lines etc., stating the results in terms of $\Omega(x,\cdot)$ only. Therefore we needed to elaborate on the geometric meaning.

6.3 The focus-directrix property

Recall that in the Euclidean case the equation

$$\operatorname{dist}(x,F) = \epsilon \cdot \operatorname{dist}(x,d),$$

where F is a point and d is a line not passing through F, determines for $0 < \epsilon < 1$ an ellipse, for $\epsilon = 1$ a parabola, and for $\epsilon > 1$ a hyperbola with a focus F and the corresponding directrix d. The following theorem gives a similar description of hyperbolic conics.

Theorem 6.4. *Let S be a non-degenerate hyperbolic conic other than a circle, horocycle, and hypercycle. Let F be a non-ideal focus of S, and let d be the corresponding directrix. Then the conic consists of all points x that satisfy the equation*

$$\delta(F,x) = \epsilon \cdot \delta(d,x) \tag{5}$$

for some positive constant ϵ, where

$$\delta(F,x) = \begin{cases} \cosh(\operatorname{dist}(x,F^\circ)) & \text{if } F \text{ is a de Sitter point,} \\ \sinh(\operatorname{dist}(x,F)) & \text{if } F \text{ is a hyperbolic point,} \end{cases}$$

$$\delta(d,x) = \begin{cases} \cosh(\operatorname{dist}(x,d^\circ)) & \text{if } d \text{ is a de Sitter line,} \\ e^{\operatorname{dist}(x,H_d)} & \text{if } d \text{ is tangent to the absolute,} \\ |\sinh(\operatorname{dist}(x,d))| & \text{if } d \text{ is a hyperbolic line.} \end{cases}$$

We need a preparatory lemma.

Lemma 6.5. *Let p be a point not on a conic Q. Then we have*

$$Q(p,p)Q(x,x) - Q(p,x)^2 \sim t_1(x) \cdot t_2(x),$$

where t_1 and t_2 are the (possibly imaginary) tangents to Q through p.

Here $f(x) \sim g(x)$ means that $f(x) = c \cdot g(x)$ for some non-zero real c.

Proof. The three conics $Q(x,x)$, $Q(p,x)^2$, and $t_1(x) \cdot t_2(x)$ belong to the same double contact pencil. Thus we have

$$\lambda Q(x,x) + \mu Q(p,x)^2 + \nu t_1(x) t_2(x) = 0.$$

For $x = p$ the third summand vanishes, which implies $\lambda = -\mu Q(p,p)$, and the lemma is proved. □

Proof of Theorem 6.4. Since the tangents through F to Ω coincide with the tangents from F to S, we have by Lemma 6.5

$$S(F,F)S(x,x) - S(F,x)^2 \sim \Omega(F,F)\Omega(x,x) - \Omega(F,x)^2,$$

and hence

$$S(x,x) \sim \Omega(F,F)\Omega(x,x) - \Omega(F,x)^2 - \lambda S(F,x)^2$$

for some $\lambda \in \mathbb{R}$. On the other hand

$$S(F,x) \sim d(x) \sim \Omega(d^\circ, x),$$

where d is the directrix corresponding to F. Hence equation $Q(x,x) = 0$ is equivalent to

$$\Omega(F,F)\Omega(x,x) - \Omega(F,x)^2 = \mu \Omega(d^\circ, x)^2,$$

for some $\mu \in \mathbb{R}$, which is in turn equivalent to

$$1 - \frac{\Omega(F,x)^2}{\Omega(F,F)\Omega(x,x)} = \nu \frac{\Omega(d^\circ,x)^2}{\Omega(x,x)}.$$

Lemma 6.2 provides metric interpretations of the expressions on the left and the right hand side. □

It can be shown that the constant in equation 5 is equal to

$$\epsilon = \sqrt{\frac{\Omega(E,E) \cdot S(F,d^\circ)}{S(E,E) \cdot \Omega(F,d^\circ)}},$$

where E is the center of the conic corresponding to the focus F, that is, the intersection point of the directrix d and the polar F° of F.

A classification of hyperbolic conics according to the nature of the (focus, directrix) pair and the value of ϵ seems to be missing in the literature. Below we describe some special cases without going into details.

Let the focus be a hyperbolic point, and the directrix be both hyperbolic. Then the equation

$$\sinh(\mathrm{dist}(x, F)) = \epsilon \cdot |\sinh(\mathrm{dist}(x, d))|$$

describes

$$\begin{cases} \text{for } 0 < \epsilon < e^{-\mathrm{dist}(d,F)} & \text{an ellipse,} \\ \text{for } \epsilon = e^{-\mathrm{dist}(d,F)} & \text{an elliptic parabola,} \\ \text{for } e^{-\mathrm{dist}(d,F)} < \epsilon < e^{\mathrm{dist}(d,F)} & \text{a semi-hyperbola,} \\ \text{for } \epsilon = e^{\mathrm{dist}(d,F)} & \text{a convex hyperbolic parabola,} \\ \text{for } \epsilon > e^{\mathrm{dist}(d,F)} & \text{a convex hyperbola.} \end{cases}$$

This can be shown by restricting the equation to the line through F perpendicular to d.

Example 6.6. For a point F and a line d not through F the locus of points satisfying

$$\mathrm{dist}(x, F) = \mathrm{dist}(x, d)$$

is a semi-hyperbola whose other focus is the pole of d, and the corresponding directrix is the polar of F, see Example 5.11. In the Beltrami–Cayley–Klein model in the unit disk, for $F = (c, 0)$ and $d = \{x = -c\}$ this semi-hyperbola is described by the equation $x = \frac{1-c^2}{4c} y^2$.

If the focus F is hyperbolic, and the directrix d is tangent to the absolute or de Sitter, then for small values of ϵ we get ellipses. As ϵ increases, the ellipses transit through elliptic parabolas to semihyperbolas, see Examples 5.11 and 5.12.

If for a semihyperbola we take its de Sitter focus, then changing the value of ϵ will transform the semihyperbola into a concave hyperbola, passing through a (long or wide) concave hyperbolic parabola or, in an exceptional case, through a horocycle.

6.4 Bifocal properties of hyperbolic conics

The following theorem is an analog of Theorem 4.8 about spherical conics.

Theorem 6.7. 1. *Let ℓ_1, ℓ_2 be a pair of lines in the hyperbolic-de Sitter plane. For a point $x \in \mathbb{H}^2$ let*

$$d_i(x) = \begin{cases} \sinh(\mathrm{dist}(x, \ell_i)) & \text{if } \ell_i \text{ is hyperbolic,} \\ \cosh(\mathrm{dist}(x, \ell_i^\circ)) & \text{if } \ell_i \text{ is de Sitter,} \\ e^{\mathrm{dist}(x, H_i)} & \text{if } \ell_i \text{ is tangent to the absolute.} \end{cases}$$

Here H_i *is a horocycle centered at the ideal point* ℓ_i°*, and the distances from a point to a line or to a horocycle are equipped with a sign. Then the locus of points that satisfy an equation of the form*

$$d_1(x) \cdot d_2(x) = c \tag{6}$$

is a hyperbolic conic, and ℓ_1, ℓ_2 *is a pair of its focal lines. Conversely, for every pair of focal lines of a hyperbolic conic, the points on the conic satisfy equation (6).*

2. *Let* p_1, p_2 *be a pair of points in the hyperbolic-de Sitter plane. For a line* ξ *in the hyperbolic plane let*

$$d_i(\xi) = \begin{cases} \sinh(\operatorname{dist}(p_i, \xi)) & \text{if } p_i \text{ is hyperbolic,} \\ e^{\operatorname{dist}(\xi, H_i)} & \text{if } p_i \text{ is ideal,} \end{cases}$$

where H_i *is a horocycle centered at* p_i*. The distances are equipped with a sign. If* p_i *is a de Sitter point, then for all oriented lines* $\xi \neq p_i^\circ$ *put*

$$d_i(\xi) = \begin{cases} \cos(\angle(\xi, p_i^\circ)) & \text{if } \xi \cap p_i^\circ \text{ is hyperbolic,} \\ \pm 1 & \text{if } \xi \cap p_i^\circ \text{ is ideal,} \\ \pm \cosh(\operatorname{dist}(\xi, p_i^\circ)) & \text{if } \xi \cap p_i^\circ \text{ is de Sitter.} \end{cases}$$

The sign convention must be chosen in such a way that $d_i(\xi)$ *depends continuously on* ξ*. Then the envelope of the lines that satisfy an equation of the form*

$$d_1(\xi) \cdot d_2(\xi) = c \tag{7}$$

is a hyperbolic conic, and p_1, p_2 *is its pair of foci. Conversely, for every pair of foci of a hyperbolic conic, the tangents to the conic satisfy equation (7).*

Proof. Let us prove the first part of the theorem.

Lemma 6.2 implies that

$$d_1(x) \cdot d_2(x) = \lambda \frac{\Omega(x, \ell_1^\circ) \cdot \Omega(x, \ell_2^\circ)}{\Omega(x, x)}$$

for some constant λ. Therefore equation (6) is equivalent to

$$\Omega(x, x) - \ell_1(x) \cdot \ell_2(x) = 0,$$

where ℓ_i are certain linear functionals on $\mathbb{R}^3$ whose kernels are the lines ℓ_i. Thus the solution set of (6) is the zero set of a quadratic form $Q = \Omega - \ell_1 \ell_2$, that is, a hyperbolic conic. If x is a (real or imaginary) intersection point of Q and Ω, then we have $\ell_1(x)\cdot\ell_2(x) = 0$, therefore the lines ℓ_1 and ℓ_2 pass through all four intersection points of Q and Ω. If some of the intersection points coincide (the conic is a parabola or a circle), then either one of the lines is an absolute tangent or both lines pass through the tangency point (this can be shown by passing to the limit). Hence ℓ_1, ℓ_2 is a pair of focal lines of the conic Q.

In the opposite direction, let Q be a conic, and let ℓ_1, ℓ_2 be its pair of focal lines. Then the conics Q, Ω, $\ell_1 \cdot \ell_2$ belong to a pencil. Therefore, up to scalar factors we have

$$Q = \Omega - \ell_1 \cdot \ell_2.$$

Thus for all points x on the conic we have $\ell_1(x) \cdot \ell_2(x) = \Omega(x,x)$, which implies $d_1(x) \cdot d_2(x) = c$ for some constant c.

The proof of the second part is similar. We have

$$d_1(\xi) \cdot d_2(\xi) = \lambda \frac{\Omega(\xi^\circ, p_1) \cdot \Omega(\xi^\circ, p_2)}{\Omega(\xi^\circ, \xi^\circ)}.$$

Therefore equation (7) is equivalent to

$$\Omega(\xi^\circ, \xi^\circ) - p_1(\xi^\circ) \cdot p_2(\xi^\circ) = 0,$$

where $p_i(x)$ is a linear functional proportional to $\Omega(p_i, x)$. This equation describes a hyperbolic-de Sitter conic; its polar is the envelope of the lines that satisfy equation (7). ☐

Example 6.8. The locus of points that satisfy

$$\sinh(\operatorname{dist}(x, \ell_1)) \cdot \sinh(\operatorname{dist}(x, \ell_2)) = c,$$

where ℓ_1, ℓ_2 are two hyperbolic lines, is

- a concave or convex hyperbola or a pair of lines, if the intersection point of f_1 and f_2 is non-ideal;
- a concave or convex hyperbolic parabola, if the intersection point of f_1 and f_2 is ideal.

Example 6.9. The locus of points that satisfy $\operatorname{dist}(x, H_1) + \operatorname{dist}(x, H_2) = c$, where H_1, H_2 are two horocycles, is a hypercycle whose ideal points are the centers of H_1 and H_2. The envelope of the lines that satisfy $\operatorname{dist}(\xi, H_1) + \operatorname{dist}(\xi, H_2) = c$ is also a hypercycle.

Theorem 6.10. *Let F_1, F_2 be a pair of foci of a hyperbolic conic. Then the conic consists of all points x that satisfy an equation of the form*

$$|\delta_1(x) + \delta_2(x)| = \text{const.} \quad \textit{or} \quad |\delta_1(x) - \delta_2(x)| = \text{const.}$$

Here the functions $\delta_i(x)$ are the distances to the foci or to their polars or to the corresponding horocycles:

$$\delta_i(x) = \begin{cases} \operatorname{dist}(x, F_i) & \textit{if } F_i \textit{ is hyperbolic,} \\ \operatorname{dist}(x, F_i^\circ) & \textit{if } F_i \textit{ is de Sitter,} \\ \operatorname{dist}(x, H_i) & \textit{if } F_i \textit{ is ideal,} \end{cases}$$

where H_i is an arbitrary horocycle centered at F_i.

Proof. Let F be a focus of the conic. First assume that F does not belong to the conic (and hence does not belong to the absolute). Denote by t_1, t_2 the two absolute tangents through F (which are both real or complex conjugate to each other). By Lemma 6.5 we have

$$Q(F,F)Q(x,x) - Q(F,x)^2 \sim t_1(x) \cdot t_2(x) \sim \Omega(F,F)\Omega(x,x) - \Omega(F,x)^2,$$

which implies

$$Q(x,x) - \frac{Q(F,x)^2}{Q(F,F)} = \lambda \cdot \left(\Omega(x,x) - \frac{\Omega(F,x)^2}{\Omega(F,F)}\right)$$

for some $\lambda \in \mathbb{R}$. To compute λ, substitute $x = E$, the center conjugate to the axis through F. Since the directrix corresponding to F also goes through E (see Lemma 5.8), we have $Q(F,E) = 0$, so that $\lambda = \frac{Q(E,E)}{\Omega(E,E)}$.

With the above argument applied to a pair of foci F_1, F_2 (of which we assume that both do not lie on the conic), we obtain

$$Q(x,x) - \lambda\Omega(x,x) = \frac{Q(F_i,x)^2}{Q(F_i,F_i)} - \lambda\frac{\Omega(F_i,x)^2}{\Omega(F_i,F_i)}, \quad \text{for } i = 1, 2.$$

Introduce the linear functions $\ell_1(x), \ell_2(x)$:

$$\ell_i(x) = \begin{cases} \dfrac{\Omega(F_i,x)}{\sqrt{\Omega(F_i,F_i)}} & \text{if } F_i \text{ is de Sitter,} \\[2ex] \dfrac{\Omega(F_i,x)}{\sqrt{-\Omega(F_i,F_i)}} & \text{if } F_i \text{ is hyperbolic.} \end{cases}$$

Since the directrices $Q(F_i, x) = 0$ and the polars $\Omega(F_i, x) = 0$ of the foci meet at the center E, each of the linear functions $Q(F_i, x)$ is a linear combination of $\ell_1(x)$ and $\ell_2(x)$. Taking into account that $Q(F_i, F_i)$ has the same sign as $\Omega(F_i, F_i)$, we obtain

$$Q - \lambda\Omega = \begin{cases} (a_{i1}\ell_1 + a_{i2}\ell_2)^2 - \lambda\ell_1^2 & \text{if } F_i \text{ is de Sitter,} \\ -(a_{i1}\ell_1 + a_{i2}\ell_2)^2 + \lambda\ell_1^2 & \text{if } F_i \text{ is hyperbolic.} \end{cases}$$

We now make a case distinction.

1) Both foci are de Sitter. We have

$$(a\ell_1 + b\ell_2)^2 - \lambda\ell_1^2 = Q - \lambda\Omega = (c\ell_1 + d\ell_2)^2 - \lambda\ell_2^2.$$

Solving this we obtain $\lambda = a^2 - b^2$. Denoting $\mu = \frac{a}{b}$ we see that

$$Q \sim (\mu^2 - 1)\Omega + \ell_1^2 + \ell_2^2 + 2\mu\ell_1\ell_2.$$

Since by Lemma 6.2

$$\frac{\ell_i(x)}{\sqrt{-\Omega(x,x)}} = \sinh(\operatorname{dist}(x, F_i^\circ)) =: \sinh(\delta_i),$$

equation $Q(x, x) = 0$ is equivalent to

$$(1 - \mu^2) + \sinh^2(\delta_1(x)) + \sinh^2(\delta_2(x)) + 2\mu \sinh(\delta_1(x)) \sinh(\delta_2(x)) = 0,$$

which factors as

$$(\cosh(\delta_1(x) + \delta_2(x)) - \mu)(\cosh(\delta_1(x) - \delta_2(x)) + \mu) = 0.$$

If $\mu > 0$, then the conic is described by $|\delta_1(x) + \delta_2(x)| = c$; if $\mu < 0$, then it is described by $|\delta_1(x) - \delta_2(x)| = c$.

2) Both foci are hyperbolic. In this case we have

$$-(a\ell_1 + b\ell_2)^2 + \lambda\ell_1^2 = Q - \lambda\Omega = -(c\ell_1 + d\ell_2)^2 + \lambda\ell_2^2,$$

which again results in $\lambda = a^2 - b^2$, so that

$$Q \sim (\mu^2 - 1)\Omega - \ell_1^2 - \ell_2^2 - 2\mu\ell_1\ell_2.$$

Now by Lemma 6.2 we have

$$\frac{\ell_i(x)}{\sqrt{-\Omega(x,x)}} = -\cosh(\operatorname{dist}(x, F_i)) =: -\cosh(\delta_i).$$

Equation $Q(x,x) = 0$ is equivalent to

$$\mu^2 - 1 + \cosh^2(\delta_1) + \cosh^2(\delta_2) + 2\mu \cosh(\delta_1)\cosh(\delta_2) = 0,$$

which factors as

$$(\cosh(\delta_1 + \delta_2) + \mu)(\cosh(\delta_1 - \delta_2) + \mu) = 0.$$

If $\mu > 0$, then the conic is empty (in fact, it is the de Sitter ellipse). If $\mu < 0$ and $\mathrm{dist}(F_1, F_2) < \mathrm{arcosh}(-\mu)$, then the second factor never vanishes, and the conic is described by $\delta_1 + \delta_2 = \mathrm{arcosh}(-\mu)$. If $\mu < 0$ and $\mathrm{dist}(F_1, F_2) > \mathrm{arcosh}(-\mu)$, then the first factor does not vanish, and the conic is described by $|\delta_1 - \delta_2| = \mathrm{arcosh}(-\mu)$.

3) Focus F_1 is de Sitter, focus F_2 is hyperbolic. We have

$$(a\ell_1 + b\ell_2)^2 - \lambda\ell_1^2 = Q - \lambda\Omega = -(c\ell_1 + d\ell_2)^2 + \lambda\ell_2^2.$$

This implies $\lambda = a^2 + b^2$, and we obtain

$$Q \sim (1 + \mu^2)\Omega - \ell_1^2 + \ell_2^2 + 2\mu\ell_1\ell_2.$$

This time we have

$$\frac{\ell_1(x)}{\sqrt{-\Omega(x,x)}} = \sinh(\mathrm{dist}(x, F_1^\circ)), \qquad \frac{\ell_2(x)}{\sqrt{-\Omega(x,x)}} = -\cosh(\mathrm{dist}(x, F_2)).$$

Equation $Q(x,x) = 0$ is equivalent to

$$1 + \mu^2 + \sinh^2(\delta_1) - \cosh^2(\delta_2) + 2\mu \sinh(\delta_1)\cosh(\delta_2) = 0,$$

which factors as

$$(\sinh(\delta_1 + \delta_2) + \mu)(\sinh(\delta_1 - \delta_2) + \mu) = 0.$$

As in the previous case, only one of the factors can vanish (depending on the relation between μ and $\mathrm{dist}(F_2, F_1^\circ)$), and the conic is described by one of the equations $\delta_1 + \delta_2 = \text{const.}$ or $\delta_1 - \delta_2 = \text{const.}$

It remains to deal with the case when one or both foci are ideal. If a focus F is ideal, then we claim that

$$Q(x,x) - \lambda\Omega(x,x) = \Omega(F,x) \cdot m(x),$$

where $\lambda = \frac{Q(E,E)}{\Omega(E,E)}$ as before, and $m(x)$ is some linear function. Indeed, the choice of λ ensures that the conic $Q - \lambda\Omega$ goes through the point E. This point belongs to the line $\Omega(F, x) = 0$, which is a common tangent of Q and Ω at the point F, and hence is also a tangent of $Q - \lambda\Omega$ at F. It follows that this line is contained in the conic $Q - \lambda\Omega$, and the statement is proved.

Now, assuming that in a pair of foci (F_1, F_2) the first one is de Sitter, and the second one is ideal, denote $\ell_2(x) = \Omega(F_2, x)$. We have

$$(a\ell_1 + b\ell_2)^2 - \lambda\ell_1^2 = Q - \lambda\Omega = \ell_2 \cdot m.$$

It follows that $\lambda = a^2$, so that

$$Q \sim \mu^2\Omega + 2\mu\ell_1\ell_2 + \ell_2^2,$$

where $\mu = \frac{a}{b}$. By Lemma 6.2, we have

$$\frac{\ell_1(x)}{\sqrt{-\Omega(x,x)}} = \sinh\delta_1, \qquad \frac{\ell_2(x)}{\sqrt{-\Omega(x,x)}} = -e^{\delta_2},$$

hence the equation $Q(x,x) = 0$ is equivalent to

$$e^{2\delta_2} - \mu e^{\delta_2+\delta_1} + \mu e^{\delta_2-\delta_1} - \mu^2 = 0,$$

which factors as

$$(e^{\delta_2+\delta_1} + \mu)(e^{\delta_2-\delta_1} - \mu) = 0.$$

Depending on whether μ is positive or negative, the conic is described by one of the equations $\delta_1 + \delta_2 = \text{const.}$ or $\delta_1 - \delta_2 = \text{const.}$

The case of an ideal and a hyperbolic focus is similar. If both foci are ideal, then we have

$$Q - \lambda\Omega = a\ell_1 \cdot \ell_2,$$

which implies that $Q(x,x) = 0$ is equivalent to $\delta_1 + \delta_2 = \text{const.}$ □

Similarly to the spherical case (Theorem 4.6), Theorem 6.10 has a dual that deals with the angles that a tangent to the conic makes with a pair of focal lines (for ultraparallel lines, the angle becomes the length of a common perpendicular). We do not list all the possibilities here, but here is one interesting particular case. Take two rays in the hyperbolic plane starting at the same point; then the segments with the endpoints on these rays that bound a triangle of constant area envelope a branch of a convex hyperbola, see Figure 36.

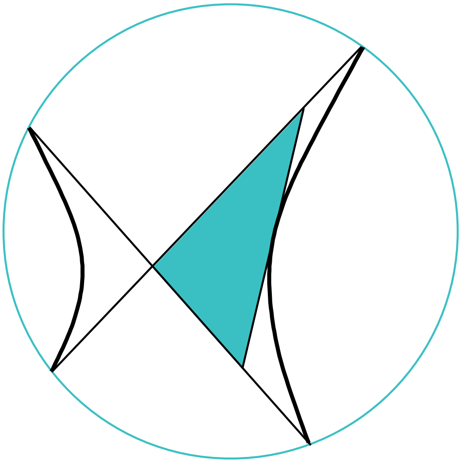


Figure 36. A pair of focal lines and a tangent to a convex hyperbola bound a triangle of constant area

A special case of Theorem 6.10 with F_1 hyperbolic, F_2 de Sitter, and zero difference of distances:

$$\operatorname{dist}(x, F_1) - \operatorname{dist}(x, F_2^\circ) = 0$$

is described in Example 6.6.

Theorem 6.10 implies that at every point of the conic the tangent to the conic forms equal angles with the gradients of the functions δ_1 and δ_2. From this we obtain an optical property of the foci, similar to the Euclidean one.

Theorem 6.11. *Let (F_1, F_2) be a pair of foci of a hyperbolic conic. Then every light ray originating from F_1 reflects from the conic in such a way that it either passes through F_2 or continues a ray originating from F_2.*

The rays originating at an ideal or de Sitter point can be defined in two ways: either as half-lines in a projective model of the hyperbolic-de Sitter plane or as the rays issued by points on a horocycle or on a line in directions orthogonal to that horocycle or a line. Figure 37 shows two examples.

A statement dual to Theorem 6.11 says that every segment tangent to the conic and having endpoints on a pair of focal lines is bisected by the point of tangency. Similar to the Euclidean and the spherical case, there is a generalization of Theorem 6.11 and of its dual, see Theorem 4.5. All these results are proved in [23] in a direct way, but the arguments are quite intricate.

Reflective properties of some hyperbolic conics were described in [19] within the Poincaré model of the hyperbolic plane.

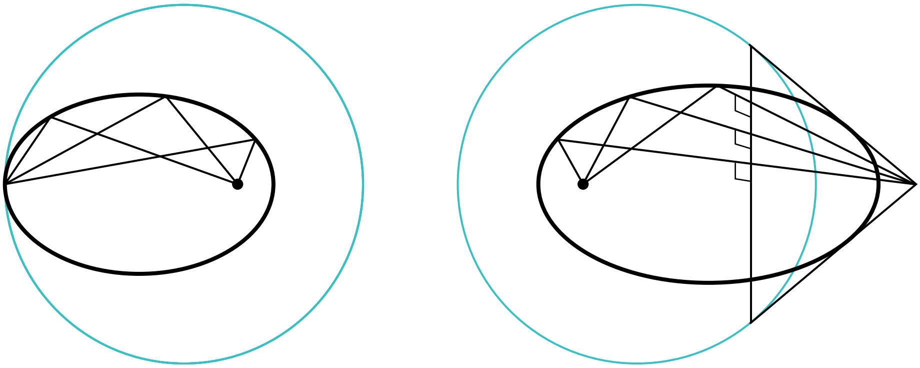


Figure 37. Optical properties of elliptic parabolas and semi-hyperbolas

References

[1] A. V. Akopyan and A. A. Zaslavsky, *Geometry of conics,* volume 26 of *Mathematical World,* Translated from the 2007 Russian original by A. Martsinkovsky, Mathematical World 26, American Mathematical Society, Providence, R.I., 2007.
MR 2359987 Zbl 1131.51008 ☞ 263, 270

[2] M. Berger, *Geometry* II, Translated from the French by M. Cole and S. Levy. Universitext, Springer-Verlag, Berlin, 1987.
MR 0882541 Zbl 0606.51001 ☞ 263, 270, 279, 299

[3] W. Blaschke, "Eine Verallgemeinerung der Theorie der konfokalen F_2," *Math. Z.* 27 (1928), no. 1, 653–668.
MR 1544932 JFM 54.0790.04 ☞ 284

[4] M. Chasles, "Résumé d'une théorie des coniques sphériques homofocales et des surfaces du second ordre homofocales," *Journal de Mathématiques Pures et Appliquées* (2) 5 (1860), 425–454.
☞ 271, 282

[5] M. Chasles and C. Graves, *Two geometrical memoirs on the general properties of cones of the second degree, and on the spherical conics,* Grant and Bolton, Dublin, 1841.
☞ 264, 284

[6] F. Dingeldey, *Encyklopädie der mathematischen Wissenschaften mit Einschluss ihrer Anwendungen,* Vol. 3, T.2, H.1: Kegelschnitte und Kegelschnittsysteme, B. G. Teubner, Leipzig, 1903, 1–160.
☞ 263

[7] F. Dingeldey and E. Fabry, *Encyclopédie des sciences mathématiques pures et appliquées,* édition française, Vol. 3: Geometrie algébrique plane. Fasc. 1: Les coniques, Gauthier-Villars, Paris, 1911, 1–160.
☞ 263

[8] T. E. Faulkner, *Projective geometry,* Oliver and Boyd, Edinburgh and London, 1960. Reprint Dover Publications, Mineola, N.Y., 2006.
MR 2311052 Zbl 1118.51002 ☞ 275

[9] F. Fillastre and A. Seppi, "Spherical, hyperbolic, and other projective geometries: convexity, duality, transitions," in *Eighteen essays in non-Euclidean geometry* (V. Alberge and A. Papadopoulos, eds.), European Mathematical Society (EMS), Zürich, 2019, Chapter 16, 321–409.
☞ 294

[10] K. Fladt, "Die allgemeine Kegelschnittsgleichung in der ebenen hyperbolischen Geometrie," *J. Reine Angew. Math.* 197 (1957), 121–139.
MR 0085546 ☞ 265

[11] K. Fladt, "Die allgemeine Kegelschnittsgleichung in der ebenen hyperbolischen Geometrie. II.", *J. Reine Angew. Math.* 199 (1958) 203–207.
MR 0095442 ☞ 265

[12] D. Fuchs and S. Tabachnikov, *Mathematical omnibus,* Thirty lectures on classic mathematics, American Mathematical Society, Providence, R.I., 2007.
MR 2350979 Zbl 1318.00004 ☞ 282

[13] G. Glaeser, H. Stachel, and B. Odehnal, *The universe of conics,* From the ancient Greeks to 21st century developments. Springer Spektrum, Berlin, 2016.
MR 3494965 Zbl 1354.51001 ☞ 263, 264, 270, 276, 293

[14] C. Graves, "On certain general properties of cones of the second degree," *Proceedings of the Royal Irish Academy* 2 (1840), 54–57.
☞ 264

[15] J. Ivory, "On the attraction of homogeneous ellipsoids," *Phil. Trans. Royal Soc. London* 99 (1809), 345–372.
☞ 282

[16] I. Izmestiev and S. Tabachnikov, “Ivory’s theorem revisited,” *J. Integrable Syst.* 2 (2017), no. 1, xyx006, 36 pp.
MR 3691968 ☞ 284

[17] F. Klein, *Vorlesungen über nicht-euklidische Geometrie.* third edition, Die Grundlehren der mathematischen Wissenschaften, 26, J. Springer, Berlin, 1928.
JFM 54.0593.01 ☞ 294

[18] L. F. Magnus, “Géométrie des surfaces courbes. Théorèmes sur l’hyperboloïde à une nappe et sur la surface conique du second ordre.” *Annales de Gergonne* 16 (1825), 33–39.
☞ 287

[19] R. Perline, “Non-Euclidean flashlights,” *Amer. Math. Monthly* 103 (1996), no. 5, 377–385.
MR 1400718 Zbl 0885.51018 ☞ 316

[20] J. G. Ratcliffe, *Foundations of hyperbolic manifolds,* second edition, Graduate Texts in Mathematics, 149, Springer, New York, 2006.
MR 2249478 Zbl 1106.51009 ☞ 306

[21] P. Samuel, *Projective geometry,* Translated from the French by S. Levy. Undergraduate Texts in Mathematics. Readings in Mathematics. Springer-Verlag, New York, 1988.
MR 0960691 Zbl 0679.51001 ☞ 266

[22] H. Stachel and J. Wallner, “Ivory’s theorem in hyperbolic spaces,” *Sibirsk. Mat. Zh.* 45 (2004), no. 4, 946–959, in Russian. English translation, *Siberian Math. J.* 45 (2004), no. 4, 785–794.
MR 2091656 Zbl 1051.51012 ☞ 282, 304

[23] W. E. Story, “On non-Euclidean properties of conics,” *Am. J. Math.* 5 (1883), 358–382.
JFM 15.0458.05 ☞ 264, 307, 316

[24] G. S. Sykes and B. Peirce, “Spherical conics,” *Proceedings of the American Academy of Arts and Sciences* 13 (1877), 375–395.
☞ 264

[25] A. Thimm, “Integrabilität beim geodätischen Fluß,” Beiträge zur Differentialgeometrie, Heft 2, Bonner Math. Schriften, 103, Univ. Bonn, Bonn, 1978.
MR 0520181 Zbl 0412.58023 ☞ 284

[26] W. P. Thurston, *Three-dimensional geometry and topology,* Vol. 1, edited by S. Levy, Princeton Mathematical Series 35, Princeton University Press, Princeton, N.J., 1997.
MR 1435975 Zbl 0873.57001 ☞ 306

[27] I. Todhunter, *A history of the mathematical theories of attraction and the figure of the earth. From the time of Newton to that of Laplace,* in two volumes, reprint of the 1873 original, Cambridge University Press, Cambridge, 2015.
Zbl 1320.85004 ☞ 282

Chapter 16

Spherical, hyperbolic, and other projective geometries: convexity, duality, transitions

François Fillastre and Andrea Seppi

Contents

1. Introduction 321
2. Model spaces 323
3. Degenerate cases 341
4. Geometric transition 359
5. Connection and volume form 372
6. Geometry of surfaces in 3-dimensional spaces 387

References 401

1 Introduction

Since the end of the 19th century, and after the works of F. Klein and H. Poincaré, it is well known that models of elliptic geometry[1] and hyperbolic geometry can be given using projective geometry, and that Euclidean geometry can be seen as a "limit" of both geometries. (We refer to [1] and [2] for historical aspects.) Then, all the geometries that can be obtained in this way (roughly speaking by defining an "absolute," which is the projective quotient of the isotropic cone of a quadratic form) were classified, see [55]. Some of these geometries had a rich development, most remarkably after the work of W. P. Thurston at the end of the 1970s, see [65] and [66], which gave rise to a highly prosperous understanding of three-dimensional hyperbolic geometry

---

[1] The elliptic space is the standard round sphere quotiented by the antipodal map.

(see among others [11], [13], [67], and [4]). On the other hand, the seminal work of G. Mess ([46], see also [6]) of 1990 motivated the study in dimension $(2 + 1)$ of the Lorentzian geometries of Minkowski, de Sitter and anti-de Sitter spaces (these spaces are known under different names in the realm of projective geometries, see [55, p. 375] or [64]), which in higher dimensions have attracted interest for a long time in mathematical physics and more precisely in General Relativity, see [49], [10], and [36]. These geometries are close relatives of hyperbolic geometry and in fact Mess outlined their strong relation with Teichmüller theory, thus giving rise to new directions of development which are still very active [6], [15], [17], [16], [8], [9], [19], [32], [62], [61], [22], and [18].

Moreover, some degenerate spaces appear naturally in the picture, namely co-Euclidean space (the space of hyperplanes of Euclidean space), and co-Minkowski space (that we will restrict to the space of space-like hyperplanes of Minkowski space), first because of duality reasons, and second because they appear as limits of degeneration of classical spaces. In fact, co-Minkowski space recently reacquired interest under the name *half-pipe geometry*, since the work of J. Danciger [26] and [27], which is very related to the idea of geometric transition [37], [25], [42], [28], [29], [20], and [21].

The purpose of this paper is to provide a survey of the properties of these spaces, especially in dimensions 2 and 3, from the point of view of projective geometry. Even with this perspective, the paper does not aim to be an exhaustive treatment. Instead it is focused on the aspects which concern convex subsets and their duality, degeneration of geometries and some properties of surfaces in three-dimensional spaces. The presentation is intended to be elementary, hence containing no proofs of deep theorems, but trying to proceed by accessible observations and elementary proofs. Apart from some constructions in Sections 5 and 6, which have been obtained in [60], there is no claim of originality in the presented results. On the other hand, in this paper we attempt to use a modern mathematical language, thus possibly contrasting with the point of view of several presentations of classical topics in the literature.

Hopefully, this survey will provide a unified introduction to the aforementioned geometries, and at the same time it might fill to some extent the lack of references that, in the opinion of the authors, surrounds the differential-geometric understanding of Minkowski, de Sitter and anti-de Sitter geometries, quite differently, for instance, from the case of hyperbolic geometry where a large number of textbooks appeared since its modern development.

2 Model spaces

2.1 Pseudo-spheres and model spaces

The set of unit vectors of three-dimensional Euclidean space, endowed with the metric induced by the ambient space, is a useful model for the round sphere $\mathbb{S}^2$. Indeed in this model,

- at a point $x \in \mathbb{S}^2$, the outward unit normal vector is the vector represented by x itself;
- the isometries of $\mathbb{S}^2$ are the restrictions of transformations in $\mathrm{O}(3)$, namely of linear isometries of Euclidean space;
- the geodesics are the intersections of $\mathbb{S}^2$ with linear planes, hence are great circles.[2]

Therefore, using this model, it is easy to show that

- $\mathbb{S}^2$ is a smooth surface;
- the isometry group of $\mathbb{S}^2$ acts transitively on points and on orthonormal frames in the tangent spaces;
- the sectional curvature is 1 at every point.

The above properties still hold for the n-dimensional sphere $\mathbb{S}^n$, the set of unit vectors of the Euclidean space of dimension $(n+1)$, which is a smooth hypersurface of constant sectional curvature if endowed with the induced metric. Moreover, it is a simple but fundamental remark that the above properties do not use the fact that the usual scalar product on the ambient space is positive definite, but only that it is a symmetric non-degenerate bilinear form.

More precisely, if a smooth manifold M is endowed with a smooth $(0,2)$-tensor which is a non-degenerate symmetric bilinear form, then there is an associated connection, the Levi-Civita connection,[3] and the notions of curvature tensor and geodesic only depend on the connection.[4] For example, in $\mathbb{R}^{n+1}$ endowed with any non-degenerate symmetric bilinear form b, the Levi-Civita connection is given by the usual differentiation in each component, the curvature is zero, and the geodesics are affine lines, regardless of the signature of b.

---

[2] This follows by contradiction using the symmetry by reflection in the plane and the fact that an unparametrized geodesic from a given point is locally uniquely determined by its tangent vector.

[3] The notion of Levi-Civita connection is recalled in Section 5.

[4] Of course, if the tensor is not positive definite, it does not induce a distance on M, and in particular there is no notion of "distance minimization" for the geodesics.

The *isotropic cone* (or null cone) of b is the set of isotropic vectors for b:

$$\mathcal{I}(b) = \{x \in \mathbb{R}^{n+1} \mid b(x,x) = 0\}.$$

Definition 2.1. A non-empty subset $\mathcal{M}$ of $\mathbb{R}^{n+1}$ is a *pseudo-sphere* if

$$\mathcal{M} = \{x \in \mathbb{R}^{n+1} \mid b(x,x) = 1\} \quad \text{or} \quad \mathcal{M} = \{x \in \mathbb{R}^{n+1} \mid b(x,x) = -1\}.$$

for a non-degenerate symmetric bilinear form b.

Remark 2.2. Of course, by replacing b by $-b$, taking 1 or -1 in the definitions leaves the topology unchanged, but produces changes at a metric level. More concretely, an *anti-isometry* between two pseudo-Riemannian spaces[5] (N_1, g_1) and (N_2, g_2) is a diffeomorphism $f\colon N_1 \to N_2$ such that $f^*g_2 = -g_1$. For example, if $b' = -b$, the pseudo-sphere

$$\mathcal{M} = \{x \mid b(x,x) = 1\},$$

endowed with the metric induced from b, is anti-isometric to

$$\mathcal{N} = \{x \mid b'(x,x) = -1\},$$

when $\mathcal{N}$ is endowed with the metric induced from b', the anti-isometry being the identity. As sets, $\mathcal{M}$ and $\mathcal{N}$ coincide, but if the signature of b is (p,q), the induced metric on $\mathcal{M}$ has signature $(p-1,q)$, while the induced metric on $\mathcal{N}$ has signature $(q, p-1)$. In the following, given $\mathcal{M}$, the pseudo-sphere $\mathcal{N}$ obtained as above will be denoted by $\overline{\mathcal{M}}$.

Let $\mathcal{M}$ be a pseudo-sphere. The following facts can be checked similarly to the corresponding properties of the unit sphere in Euclidean space:

- at a point $x \in \mathcal{M}$, a unit normal vector is the vector represented by x itself,
- the restrictions of the transformations of $\mathrm{O}(p,q)$, namely the linear isometries of b, are isometries[6] of $\mathcal{M}$,
- totally geodesic submanifolds are the intersections of $\mathcal{M}$ with vector subspaces.

---

[5] A smooth manifold is *pseudo-Riemannian* if it is endowed with a non-degenerate smooth $(0,2)$-tensor. If the tensor is positive definite at each point, then the manifold is *Riemannian*. If the tensor has a negative direction, but no more than one linearly independent negative direction at each point, then the manifold is *Lorentzian*.

[6] The isometry group of $\mathcal{M}$ is bigger than $\mathrm{O}(p,q)$ if $\mathcal{M}$ is not connected.

It follows straightforwardly from the above properties that

- $\mathcal{M}$ is a smooth hypersurface,
- the isometry group of $\mathcal{M}$ acts transitively on points and on orthonormal frames in the tangent spaces,
- the sectional curvature is constantly equal to 1 if $\mathcal{M} = b^{-1}(1)$ and equal to -1 if[7] $\mathcal{M} = b^{-1}(-1)$.

For $n \geq 2$, the manifold $b^{-1}(1)$ is diffeomorphic to $\mathcal{S}^{p-1} \times \mathbb{R}^q$, and $b^{-1}(-1)$ is diffeomorphic to $\mathcal{S}^{q-1} \times \mathbb{R}^p$, so it may be disconnected. As for the remainder of this section, the reference is [49] for more details. But from the definition, if $x \in \mathcal{M}$, then also $-x \in \mathcal{M}$, hence we will be mainly interested in the projective quotient of $\mathcal{M}$. As the action of the antipodal map is clearly isometric, the quotient will inherit in an obvious way many of the properties of $\mathcal{M}$.

Definition 2.3. A *model space* $\mathbb{M}$ is a space $\mathbb{M} := \mathcal{M}/\{\pm \mathrm{Id}\}$, where $\mathcal{M}$ is a pseudo-sphere, together with the induced metric from $b|_{\mathcal{M}}$.

It is obvious that the projective quotient of $\overline{\mathcal{M}}$ is anti-isometric to $\mathbb{M}$, and will thus be denoted by $\overline{\mathbb{M}}$. Moreover, in the case where $\mathcal{M}$ is not connected, it has two connected components, which correspond under the antipodal map $x \mapsto -x$. In particular, $\mathbb{M}$ is connected.

By construction, $\mathbb{M}$ is a subset of the projective space $\mathbb{R}\mathrm{P}^n$. Indeed, topologically $\mathbb{M}$ can be defined as

$$\mathbb{M} = \mathrm{P}\{x \in \mathbb{R}^{n+1} \mid b(x,x) > 0\},$$

or the same definition with $>$ replaced by $<$, depending on the case. Hence, any choice of an affine hyperplane in $\mathbb{R}^{n+1}$ which does not contain the origin will give an affine chart of the projective space, and the image of $\mathbb{M}$ in the affine chart will be an open subset of an affine space of dimension n.

Recall that $\mathrm{PGL}(n+1,\mathbb{R})$, the group of *projective transformations* (or homographies), is the quotient of $\mathrm{GL}(n+1,\mathbb{R})$ by the non-zero scalar transformations. Even if $\mathcal{M}$ is not connected, an isometry of $\mathcal{M}$ passes to the quotient only if it is an element of $\mathrm{O}(p,q)$. Hence $\mathrm{Isom}(\mathbb{M})$, the isometry group of $\mathbb{M}$, is $\mathrm{PO}(p,q)$, the quotient of $\mathrm{O}(p,q)$ by $\{\pm \mathrm{Id}\}$ (indeed elements of $\mathrm{O}(p,q)$ have determinant equal to 1 or -1).

---

[7] This follows from the Gauss formula and the fact that the shape operator on $\mathcal{M}$ is $\pm$ the identity. The fact that the sectional curvature is constant also follows from the preceding item.

2.2 Lines and pseudo-distance

Pseudo-distance on $\mathcal{M}$. A geodesic c of $\mathcal{M}$ is the non-empty intersection of $\mathcal{M}$ with a linear plane. This intersection might not be connected. The geodesic c of $\mathcal{M}$ is said to be

- *space-like* if $b(\dot{c}, \dot{c}) > 0$,
- *light-like* (or *null*) if $b(\dot{c}, \dot{c}) = 0$,
- *time-like* if $b(\dot{c}, \dot{c}) < 0$.

An anti-isometry sends space-like lines onto time-like lines. For example, assume that the restriction of b to a plane P of coordinates (x_1, x_2) has the form $x_1^2 + x_2^2$. Then the curve

$$\{x \in P \mid b(x, x) = 1\} = \{x_1^2 + x_2^2 = 1\}$$

is a space-like geodesic. On the other hand, $b' = -b$ has the form $-x_1^2 - x_2^2$ and thus

$$\{x \in P \mid b'(x, x) = -1\} = \{x_1^2 + x_2^2 = 1\}$$

is time-like. Actually, up to a global isometry of the ambient space, a non-light-like geodesic of $\mathcal{M}$ can be written in one of the following forms:

(a1) $\{x \in P \mid b(x, x) = 1\}$, where $b|_P(x, x) = x_1^2 + x_2^2$;

(a2) $\{x \in P \mid b(x, x) = -1\}$, where $b|_P(x, x) = -x_1^2 - x_2^2$;

(a3) $\{x \in P \mid b(x, x) = 1\}$, where $b|_P(x, x) = x_1^2 - x_2^2$;

(a4) $\{x \in P \mid b(x, x) = -1\}$, where $b|_P(x, x) = x_1^2 - x_2^2$.

Up to a change of coordinates, a non-light-like geodesic is a circle or a branch of hyperbola. If the geodesic is light-like, it is an affine line in the ambient space. It follows from these descriptions that:

Fact 2.4. *Every pseudo-sphere $\mathcal{M}$ is* ***geodesically complete****: each maximal (for the inclusion) geodesic is defined on the entire real line.*

Let x, y be on a connected component of a geodesic c of $\mathcal{M}$, and let us parametrize by $\gamma\colon [0, 1] \to c$ a subsegment of c such that $\gamma(0) = x$ and $\gamma(1) = y$. The *pseudo-distance* is

$$\tilde{d}(x, y) = \int_0^1 \sqrt{|b(\dot{\gamma}(t), \dot{\gamma}(t))|}dt. \tag{1}$$

If x and y lie on a circle, there are actually two choices for the subsegment of c joining them. We choose the one which gives the smaller distance. In particular, in this case $\tilde{d}(x,y) \in [0,\pi]$. Moreover, if c is light-like, then $\tilde{d}(x,y) = 0$. We have $\tilde{d}(x,x) = 0$ and $\tilde{d}(x,y) = \tilde{d}(y,x)$, but in general $\tilde{d}$ is not a distance.[8]

Considering the cases in (a1)–(a4), explicit parameterizations of the circle or the hyperbola give:

(b1) $b(x,y) = \cos(\tilde{d}(x,y))$;

(b2) $-b(x,y) = \cos(\tilde{d}(x,y))$;

(b3) $b(x,y) = \cosh(\tilde{d}(x,y))$;

(b4) $-b(x,y) = \cosh(\tilde{d}(x,y))$.

Projective distance on $\mathbb{M}$. We are mainly interested in non-parametrized geodesics of $\mathbb{M}$.

Definition 2.5. A *line* (resp. *plane*, *hyperplane*) of $\mathbb{M}$ is the projective quotient of a non-empty intersection of $\mathcal{M}$ with a 2-dimensional (resp. 3-dimensional, n-dimensional) vector space.

A straightforward property is that the image of a line (resp. plane, hyperplane) of $\mathbb{M}$ in an affine chart is the intersection of the image of $\mathbb{M}$ with an affine line (resp. plane, hyperplane) of $\mathbb{R}^n$.

We say that a line of $\mathbb{M}$ is *space-like* (respectively *time-like*, *light-like*) if its lift is a space-like (resp. time-like, light-like) geodesic of $\mathcal{M}$. But also lines can be classified with respect to the quadric given by the projective quotient of the isotropic cone.

Definition 2.6. The projective quotient $\mathrm{P}\mathcal{I}(b)$ of the isotropic cone is called the *absolute*.[9]

Definition 2.7. A line of $\mathbb{M}$ is called

- *hyperbolic* if it intersects the absolute in two distinct points,
- *parabolic* if it intersects the absolute in one point,
- *elliptic* it it does not intersect the absolute.

---

[8] It is a distance when $\mathcal{M}$ is Riemannian.

[9] The absolute corresponds to what is sometimes called the *boundary at infinity* of $\mathbb{M}$. We will avoid this terminology in the present paper, because some confusion may occur from the fact that in an affine chart, the boundary at infinity may be at infinity or not. Note that an absolute is often the boundary at infinity of two model spaces.

Lemma 2.8. *Parabolic and light-like lines coincide.*

Proof. A parabolic line is the projective quotient of a plane P containing a unique line belonging to the isotropic cone of b. But a light-like geodesic of $\mathcal{M}$ is the intersection of $\mathcal{M}$ with a plane on which the restriction of b is degenerate. ☐

Recall that an anti-isometry exchanges space-like and time-like lines, but it preserves the type of a line given by Definition 2.7. Actually the lines in $\mathbb{M}$ whose lifts are described in (a1)–(a4) have the following corresponding type:

(c1) space-like elliptic,

(c2) time-like elliptic,

(c3) time-like hyperbolic,

(c4) space-like hyperbolic.

If x, y are on a line of $\mathbb{M}$, we will define a "distance" between them, which roughly speaking corresponds to the pseudo-distance between the corresponding lifts. But one has to choose the lifts carefully.

Definition 2.9. Let x, y be on a line l of $\mathbb{M}$. The *projective distance* $d(x, y)$ is

- 0 if l is parabolic,
- $\tilde{d}(\tilde{x}, \tilde{y})$ if l is hyperbolic, and $\tilde{x}, \tilde{y}$ are lifts of x and y on the same branch of hyperbola,
- $\tilde{d}(\tilde{x}, \tilde{y})$ if l is elliptic, and $\tilde{x}, \tilde{y}$ are lifts of x and y such that $\tilde{d}(\tilde{x}, \tilde{y})$ is the smallest.[10]

With the help of the cross-ratio, one can give a more direct definition of the projective distance. Recall that if four points x, y, q, p are given on the projective line $\mathbb{R}\mathrm{P}^1 \cong \mathbb{R} \cup \{\infty\}$, then there exists a unique projective transformation h with $h(x) = \infty$, $h(y) = 0$ and $h(q) = 1$. Then the *cross-ratio* can be defined as

$$[x, y, q, p] = h(p) \in \mathbb{R} \cup \{\infty\}.$$

Of course, $[x, y, q, p] = \infty$ if $p = x$, $[x, y, q, p] = 0$ if $p = y$ and $[x, y, q, p] = 1$ if $p = q$. More explicitly, a formula is

$$[x, y, q, p] = \frac{x - q}{y - q}\frac{y - p}{x - p}.$$

Note that this last definition extends to complex numbers, i.e. the definition of the cross-ratio extends to $\mathbb{C}\mathrm{P}^1$.

---

[10] In particular, in this case $d(x, y) \in [0, \frac{\pi}{2}]$.

Lemma 2.10. *Given two distinct points $x, y \in \mathbb{M}$ on a line l, then l intersects the absolute at two points* I, J (*may be not distinct, may be in* $\mathbb{C}\,\mathrm{P}^1$), *and*

$$d(x, y) = \left|\frac{1}{2}\ln[x, y, \mathrm{I}, \mathrm{J}]\right|. \tag{2}$$

Here $\ln$ *denotes the branch of the complex logarithm* $\ln\colon \mathbb{C} \setminus \{0\} \to \{z \in \mathbb{C}\colon \mathrm{Im}(z) \in (-\pi, \pi]\}$.

Proof. If l is parabolic, then $\mathrm{I} = \mathrm{J}$, $[x, y, \mathrm{I}, \mathrm{J}] = 1$ and the result follows.

Let us consider the case when l is elliptic. A lift of the line containing x and y is a connected component of the intersection of a 2-plane P with $\mathcal{M}$. The restriction of the bilinear form b to the plane P has necessarily signature $(+,+)$ or $(-,-)$, otherwise P would meet the isotropic cone. Without loss of generality, let us consider the case $(+,+)$, and let us introduce coordinates such that

$$b(x, x) = x_1^2 + x_2^2$$

and $c = P \cap \mathcal{M} = \{x \in P \mid b(x, x) = 1\}$. In the complex plane, the equation $x_1^2 + x_2^2 = 0$ is solved by the two complex lines[11]

$$\mathrm{I} = \begin{bmatrix} i \\ 1 \end{bmatrix} \text{ and } \mathrm{J} = \begin{bmatrix} -i \\ 1 \end{bmatrix}.$$

For more about the points I, J see e.g. [53].

Let us parametrize c as $(\cos t, \sin t)$, and let us choose t_x and t_y such that $d(x, y) = |t_x - t_y|$. In the affine chart $x_2 = 1$, the line I (resp. J, x, y) is represented by the point i (resp. $-i$, $\operatorname{cotan} t_x$, $\operatorname{cotan} t_y$). Then we compute

$$\begin{aligned}
\frac{1}{2}\ln[x, y, \mathrm{I}, \mathrm{J}] &= \frac{1}{2}\ln[\operatorname{cotan} t_x, \operatorname{cotan} t_y, i, -i] \\
&= \frac{1}{2}\ln\left(\frac{i \operatorname{cotan} t_x + 1}{i \operatorname{cotan} t_y + 1}\,\frac{i \operatorname{cotan} t_y - 1}{i \operatorname{cotan} t_x - 1}\right) \\
&= \frac{1}{2}\ln\left(\frac{i \operatorname{cotan} t_x + 1}{i \operatorname{cotan} t_x - 1}\right) - \frac{1}{2}\ln\left(\frac{i \operatorname{cotan} t_y + 1}{i \operatorname{cotan} t_y - 1}\right) \\
&= \operatorname{arcoth}(i \operatorname{cotan} t_x) - \operatorname{arcoth}(i \operatorname{cotan} t_y) \\
&= i t_x - i t_y \\
&= i(t_x - t_y),
\end{aligned}$$

---

[11] In the text, we will use brackets to designate a projective equivalence class.

where the last equation follows from $\cos(x) = \cosh(ix)$, $\sin(x) = \left(\frac{1}{i}\right)\sinh(ix)$, therefore $i \operatorname{cotan} x = \operatorname{arcoth}(ix)$. This shows that, if l is elliptic, then the expression inside the modulus of (2) is purely imaginary. Taking the modulus we obtain

$$\left|\frac{1}{2}\ln[x, y, \mathrm{I}, \mathrm{J}]\right| = d(x, y).$$

If l is hyperbolic, it intersects the absolute in two distinct points I, J, such that I and J do not separate x and y. Similarly to the previous case, one can assume $b(x, x) = x_1^2 - x_2^2$ on P and $c = P \cap \mathcal{M} = \{x \in P \mid b(x, x) = -1\}$. Picking the lifts of x and y on the same branch of the hyperbola of c, which is parameterized by $(\cosh t, \sinh t)$, one sees that in the affine line $x_2 = 1$ the line I (resp. J, x, y) is represented by the point -1 (resp. 1, $\tanh t_x$, $\tanh t_y$). Hence by an analogous computation,

$$\begin{aligned}\frac{1}{2}\ln[x, y, \mathrm{I}, \mathrm{J}] &= \frac{1}{2}\ln[\tanh t_x, \tanh t_y, -1, 1]\\ &= \frac{1}{2}\ln\left(\frac{\tanh t_x + 1}{\tanh t_x - 1}\right) - \frac{1}{2}\ln\left(\frac{\tanh t_y + 1}{\tanh t_y - 1}\right)\\ &= t_x - t_y.\end{aligned}$$

In this case $[x, y, \mathrm{I}, \mathrm{J}]$ is a real number, and the formula holds in a way analogous to the elliptic case. ☐

2.3 Usual model spaces

We will often consider the symmetric bilinear form b on $\mathbb{R}^{n+1}$ in a standard form, i.e. such that the standard basis of $\mathbb{R}^{n+1}$ is an orthonormal basis for b, and if the signature of b is (p, q), then the quadratic form is positive on the first p vectors of the standard basis, and negative on the other q vectors. In this case, we will denote the bilinear form by $b_{p,q}$.

Euclidean and Elliptic spaces. In our notation $\mathbb{R}^{n+1}$ is the real coordinate space of dimension $(n + 1)$, and we will use the notation $\mathbb{E}^{n+1}$ for the Euclidean space of dimension $(n + 1)$, that is, $\mathbb{R}^{n+1}$ endowed with $b_{n+1,0}$. The *elliptic space* $\mathbb{E}\mathbb{l}\mathbb{l}^n$ is the projective space of dimension n endowed with the spherical metric inherited from $\mathbb{S}^n$, namely

$$\mathbb{E}\mathbb{l}\mathbb{l}^n = \{x \in \mathbb{R}^{n+1} \mid b_{n+1,0}(x) = 1\}/\{\pm \mathrm{Id}\}.$$

It is a compact Riemannian manifold of dimension n. Any affine chart corresponds to a central projection of the sphere onto a hyperplane. All its lines are space-like elliptic, in particular the distance induced by the Riemannian structure is given by the projective distance (2).

Minkowski, Hyperbolic and de Sitter spaces. The space $\mathbb{R}^{n+1}$ endowed with $b_{n,1}$ is called the *Minkowski space* and denoted by $\mathbb{Min}^{n+1}$. An affine hyperplane Q of direction P in $\mathbb{Min}^{n+1}$ is called

- *space-like* if $(b_{n,1})|_P$ is positive definite;
- *light-like* or *null* if $(b_{n,1})|_P$ is degenerate;
- *time-like* if $(b_{n,1})|_P$ has signature $(n-1, 1)$.

(See Figure 1 for the $n = 2$ case.)

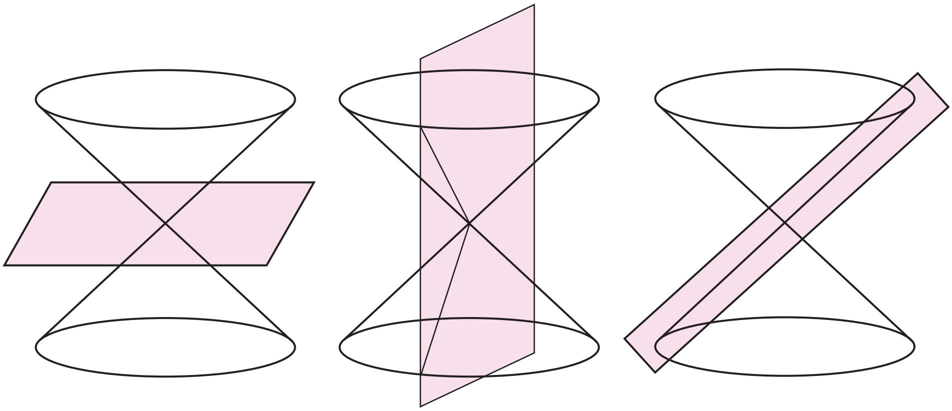


Figure 1. A plane in $\mathbb{Min}^3$ is space-like (resp. time-like, light like) if its intersection with the isotropic cone of $b_{2,1}$ is reduced to a point (resp. two lines, a single line)

The *hyperbolic space* is

$$\mathbb{H}^n = \mathcal{H}^n/\{\pm \operatorname{Id}\},$$

with

$$\mathcal{H}^n = \{x \in \mathbb{R}^{n+1} \mid b_{n,1}(x) = -1\},$$

endowed with the induced metric. The hyperbolic space is a simply connected complete Riemannian surface of sectional curvature -1. All its line are space-like hyperbolic, in particular the distance induced by the Riemannian structure is given by the projective distance (2).

The other pseudo-sphere of Minkowski space is the *de Sitter space*

$$\mathrm{d}\mathbb{S}^n = \{x \in \mathbb{R}^{n+1} \mid b_{n,1}(x) = 1\}/\{\pm \mathrm{Id}\}$$

endowed with the induced metric. On any tangent plane to $\mathrm{d}\mathbb{S}^n$, the restriction of $b_{n,1}$ has Lorentzian signature, i.e. it is non-degenerate with signature $(n-1, 1)$. De Sitter space has constant curvature 1, and is homeomorphic to $\mathbb{S}^{n-1} \times \mathbb{R}$.

Let us focus our attention to the case $n = 2$. The de Sitter plane $\mathrm{d}\mathbb{S}^2$ has space-like elliptic, null and time-like hyperbolic lines. This is more easily seen by taking affine models. The most common model is $\{x_3 = 1\}$. Then $\mathbb{H}^2$ is the unit open disc, and $\mathrm{d}\mathbb{S}^2$ is the complement of the closed disc in the plane, see Figure 2.[12]

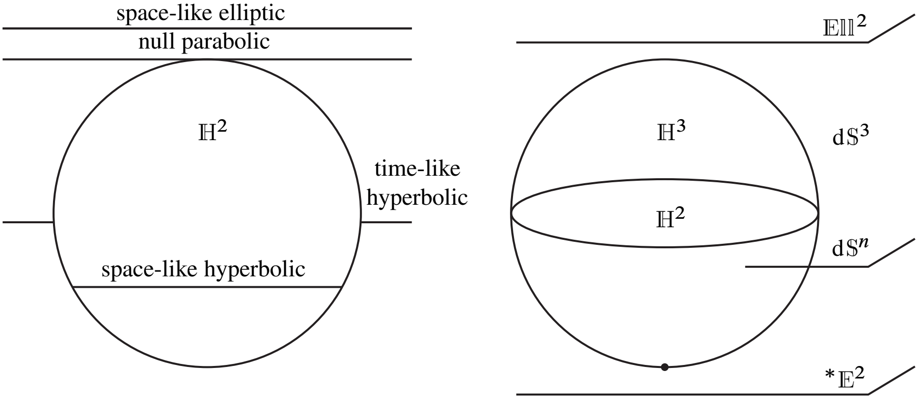


Figure 2. On the left: types of lines in an affine model of the hyperbolic and de Sitter planes. On the right: types of planes in an affine model of the hyperbolic and de Sitter spaces.

By any affine transformation, any ellipse instead of the circle gives another affine model. (These affine transformations are given by the choice of any space-like affine plane in $\mathbb{M}\mathrm{in}^3$ which does not contain the origin.) Taking as affine chart a time-like or a light-like affine plane of $\mathbb{M}\mathrm{in}^3$, we obtain other models of the hyperbolic and de Sitter planes:

Fact 2.11. *The convex side of any conic of the plane* (*i.e. a parabola, ellipse or a hyperbola*) *endowed with the projective distance is an affine model of* $\mathbb{H}^2$ *or* $\overline{\mathbb{H}^2}$. *The other side is an affine model of* $\mathrm{d}\mathbb{S}^2$ *or* $\overline{\mathrm{d}\mathbb{S}^2}$.

---

[12] This is the famous *Klein model* for the hyperbolic plane.

In dimension 3, we obtain:

Fact 2.12. *Any convex side of a non-degenerate quadric of* $\mathbb{R}^3$ *(i.e. an ellipsoid, an elliptic paraboloid or a two-sheeted hyperboloid) endowed with the projective distance is a model for* $\mathbb{H}^3$ *or* $\overline{\mathbb{H}^3}$. *The other side is a model for* $\mathrm{d}\mathbb{S}^3$ *or* $\overline{\mathrm{d}\mathbb{S}^3}$.

It is then immediate that a plane of $\mathbb{H}^3$ is a hyperbolic plane, and that a plane of $\mathrm{d}\mathbb{S}^3$ may be isometric to $\mathrm{d}\mathbb{S}^2$, to the elliptic plane $\mathbb{Ell}^2$, or to the co-Euclidean plane $^*\mathbb{E}^2$ (see 3.5), see Figure 2.

Anti-de Sitter space. By considering the non-degenerate symmetric bilinear form $b_{n-1,2}$ on $\mathbb{R}^{n+1}$, for $n \geq 2$, we define the *anti-de Sitter space* as

$$\mathbb{A}\mathrm{d}\mathbb{S}^n = \mathcal{A}d\mathcal{S}^n/\{\pm \mathrm{Id}\},$$

with

$$\mathcal{A}d\mathcal{S}^n = \{x \in \mathbb{R}^{n+1} \mid b_{n-1,2}(x) = -1\}$$

endowed with the metric induced from $b_{n-1,2}$. Anti-de Sitter space is a Lorentzian manifold of curvature -1. It is not simply connected as it contains a closed time-like geodesic.

Remark 2.13. It is readily seen that $\mathbb{A}\mathrm{d}\mathbb{S}^2$ is anti-isometric to $\mathrm{d}\mathbb{S}^2$. This is not the case for $n > 2$.

Let us now focus on the $n = 3$ case. First note that $\mathcal{A}d\mathcal{S}^3$ is anti-isometric to $b_{2,2}^{-1}(1)$. Up to affine transformations, $\mathbb{A}\mathrm{d}\mathbb{S}^3$ has two models described as follows:

Fact 2.14. *Any part of* $\mathbb{R}^3$ *delimited by a double-ruled quadric (i.e. a hyperbolic paraboloid or a one-sheeted hyperboloid) endowed with the projective distance is a model of* $\mathbb{A}\mathrm{d}\mathbb{S}^3$ *or* $\overline{\mathbb{A}\mathrm{d}\mathbb{S}^3}$.

A plane in $\mathbb{A}\mathrm{d}\mathbb{S}^3$ can be isometric to $\mathbb{H}^2$, to the co-Minkowski plane $^*\mathbb{Min}^2$ (cf. Section 3.5), or to $\mathbb{A}\mathrm{d}\mathbb{S}^2$, see Figure 3.

2.4 Convex sets and duality

Dual cones. A non-empty subset K of $\mathbb{R}^{n+1}$ is *convex* if it contains the segment between any two of its points (note that this notion only uses the affine structure of $\mathbb{R}^{n+1}$). A convex cone $\mathcal{C}$ of $\mathbb{R}^n$ is a convex set such that $\lambda\mathcal{C} \subset \mathcal{C}$ for any $\lambda > 0$.

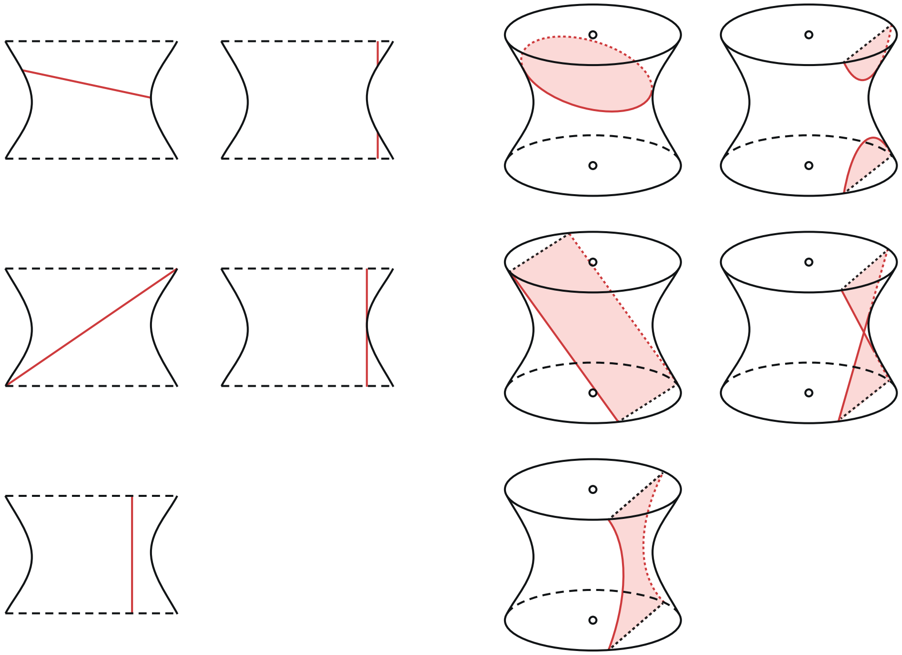


Figure 3. On the left-hand side, some non-affinely equivalent lines in an affine model of $\mathbb{A}\mathrm{d}\mathbb{S}^2$. First row on the left: space-like hyperbolic, second row: light-like parabolic, third row: time-like elliptic. On the right-hand side, some non-affinely equivalent planes in an affine model of $\mathbb{A}\mathrm{d}\mathbb{S}^3$. First row: hyperbolic, second row: co-Minkowski, third row: $\mathbb{A}\mathrm{d}\mathbb{S}^2$.

We will also suppose that $\mathcal{C}$ is pointed: the only linear subspace it contains is $\{0\}$. In general, we will also assume implicitly that the cone is closed.

The dual $(\mathbb{R}^{n+1})^*$ of $\mathbb{R}^{n+1}$ is the set of linear forms on $\mathbb{R}^{n+1}$, and is naturally endowed with a vector space structure of dimension $(n+1)$. Note that the notion of convexity also holds in $(\mathbb{R}^{n+1})^*$. Given a convex cone $\mathcal{C}$ in $\mathbb{R}^n$, its *dual* is

$$\mathcal{C}^* = \{x^* \in (\mathbb{R}^n)^* \mid x^*(y) \leq 0, \text{ for all } y \in \mathcal{C}\}. \tag{3}$$

It is readily seen that the dual of a convex cone is a convex cone, and that

$$A \subset B \implies B^* \subset A^*. \tag{4}$$

Recall that a *support space* of a closed convex set K is a half-space containing K and bounded by an affine hyperplane H. If moreover $K \cap H \neq \emptyset$, H is a *support plane* of K. A convex set is also the intersection of its support spaces. If $K = \mathcal{C}$ is a convex cone, then its support planes are linear hyperplanes. Now if x^* is a non trivial linear form on $\mathbb{R}^{n+1}$, then its kernel is a linear hyperplane, and $\{y \mid x^*(y) \leq 0\}$ is a half-space bounded by this hyperplane. So $\mathcal{C}^*$ can be interpreted as the set of support planes of $\mathcal{C}$. As a convex set is the intersection of its support spaces,

$$\mathcal{C} = \{x \in \mathbb{R}^n \mid y^*(x) \le 0, \text{ for all } y^* \in \mathcal{C}^*\}. \tag{5}$$

From the formal duality between the definitions (3) and (5), it is readily seen that

$$(\mathcal{C}^*)^* = \mathcal{C}. \tag{6}$$

If $\mathbb{R}^{n+1}$ is endowed with a non-degenerate symmetric bilinear form b, then b induces an isomorphism between $\mathbb{R}^{n+1}$ and $(\mathbb{R}^{n+1})^*$. Once b, and thus the isomorphism, are fixed, we will still denote by $\mathcal{C}^*$ the image of $\mathcal{C}^*$ in $\mathbb{R}^{n+1}$:

$$\mathcal{C}^* = \{x \in \mathbb{R}^n \mid b(x, y) \le 0, \text{ for all } y \in \mathcal{C}\}.$$

See Figure 4 for an example with $b = b_{1,1}$. Let us denote by C the projective quotient of $\lambda\mathcal{C}$, for all $\lambda \in \mathbb{R}$, i.e.

$$\mathrm{C} = (\mathcal{C} \cup -\mathcal{C})/\{\pm \mathrm{Id}\}.$$

Remark 2.15. We could have defined as a dual for $\mathcal{C}$ the convex set

$$\mathcal{C}^{*\prime} = \{x \in \mathbb{R}^n \mid b(x, y) \ge 0, \text{ for all } y \in \mathcal{C}\},$$

as both definitions agree in the projective quotient: $\mathrm{C}^* = \mathrm{C}^{*\prime}$.

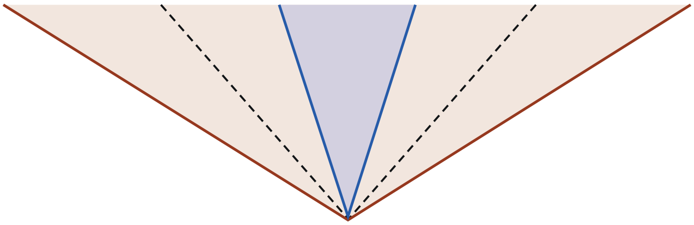


Figure 4. Dual cones in the Minkowski plane

Convex sets in model spaces. Let $\mathcal{M}$ be a pseudo-sphere, $\mathcal{C}$ be a convex cone, and suppose that $\mathcal{M} \cap \mathcal{C}$ is non-empty. Recall that $\mathbb{M}$ is the projective quotient of $\mathcal{M}$, that is, $\mathbb{M} := \mathcal{M}/\{\pm \mathrm{Id}\}$.

Definition 2.16. A *convex set* of $\mathbb{M}$ is the intersection $\mathbb{M} \cap \mathrm{C}$.

The image of a convex cone in an affine chart may not be an affine convex set. For example the intersection of $\{z^2 \leq x^2 + y^2\}$ with a vertical plane. But the image of a convex cone in an affine chart is an affine convex set if the vector hyperplane parallel to the hyperplane which defines the affine chart meets the cone only at $\{0\}$. Moreover in an affine chart, even if C is convex, $\mathbb{M} \cap \mathrm{C}$ is not necessarily an affine convex set (the image of $\mathbb{M}$ in the affine chart may be non convex).

Also, a convex set of $\mathbb{M}$ is not necessarily a geodesically convex subset of $\mathbb{M}$. For example, take any convex cone in $\mathbb{R}^3$ which contains the isotropic cone of $b_{2,1}$ in its interior, and is on one side of the horizontal plane. In the affine chart given by $\{x_3 = 1\}$, this gives a convex set of $\mathrm{d}\mathbb{S}^2$, which is an affine convex set. But it contains the absolute in its interior, so it is not a geodesically convex set of $\mathrm{d}\mathbb{S}^2$ (see the big triangle in the left-hand side of Figure 5).

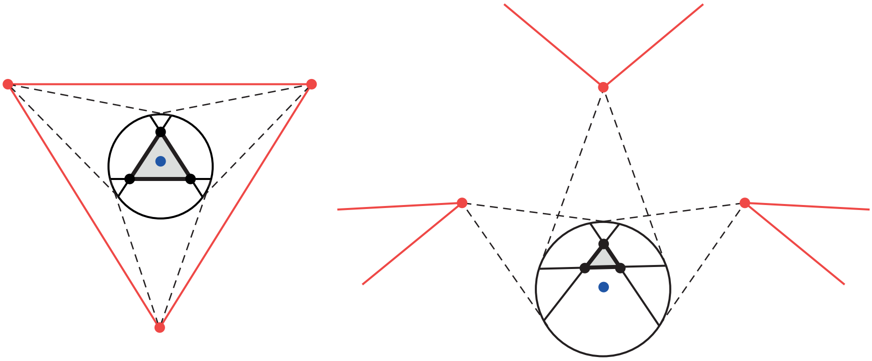


Figure 5. Triangles in $\mathrm{d}\mathbb{S}^2$ dual to triangles in $\mathbb{H}^2$ in an affine chart. Note that on the right-hand side, the dual triangle is not the convex hull in the affine chart of the vertices.

Duality. The notion of duality for convex sets follows easily from the one for convex cones. Let us focus on the more relevant cases. Let K be a convex set in $\mathbb{H}^n$, defined by a convex cone $\mathcal{C}(K)$. Using the bilinear form $b_{n,1}$ to identify the ambient space $\mathbb{R}^{n+1}$ with its dual, the support planes of $\mathcal{C}(K)$ are time-like or light-like, and the boundary of $\mathcal{C}(K)^*$ is made of space-like or light-like lines, and hence its intersection K^* with $\mathrm{d}\mathbb{S}^n$ is not empty. The set K^* is a convex set of de Sitter space, the *dual* of K. As the lines of $\mathcal{C}(K)$ are orthogonal to the support planes of $\mathcal{C}(K)^*$ by construction, K^* is *space-like*: it has only space-like or light-like support planes. Conversely, if K is a space-like convex set in $\mathrm{d}\mathbb{S}^n$, then we can define in the same way its dual convex set K^*, which is a convex set of $\mathbb{H}^n$.

If K is a convex set in $\mathbb{E}\mathbb{ll}^n$, we obtain similarly a dual convex set K^* which is in $\mathbb{E}\mathbb{ll}^n$. Analogously, the dual of a space-like convex set K in $\mathbb{A}\mathrm{d}\mathbb{S}^3$ is a space-like convex set K^* in $\mathbb{A}\mathrm{d}\mathbb{S}^3$. In any case, we have $(K^*)^* = K$.

The duality is also a duality between points and hyperplanes. For $x \in \mathcal{M} = b^{-1}(1)$, x^* is the intersection (if non-empty) of $\mathcal{N} = b^{-1}(-1)$ with the hyperplane of $\mathbb{R}^{n+1}$ orthogonal to x for b.

This duality points/hyperplane can also be seen in an affine manner. Let us begin with the hyperbolic case. A hyperplane in $\mathbb{H}^n$ meets the absolute along a topological sphere S of dimension $n-2$. Then the point dual to the hyperplane is the apex of the cone formed by the lines tangent to the absolute along S. This is the usual notion of polarity transformation with respect to a (proper) quadric, which is an affine notion, see Figure 6 and Figure 7.

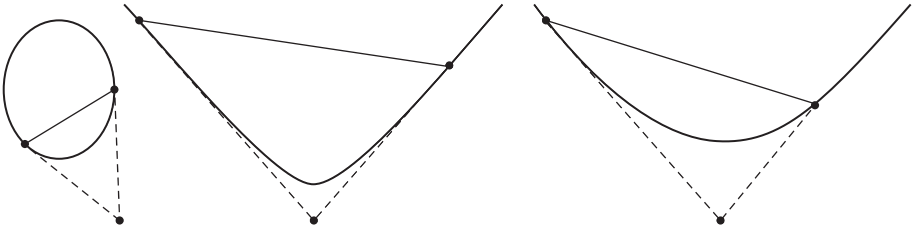


Figure 6. The dual point of a line in different affine models of the hyperbolic plane

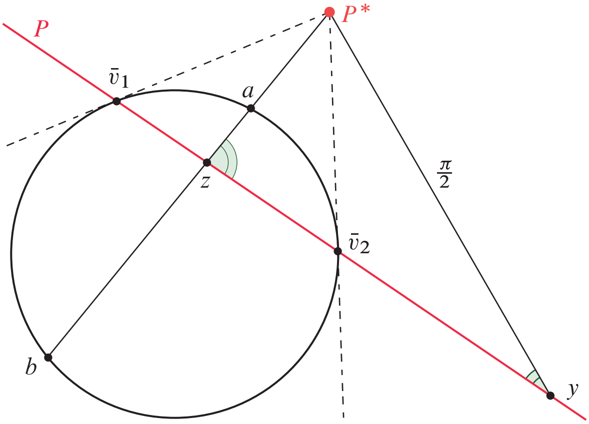


Figure 7. The hyperplane containing P meets the absolute along a hypersphere S. The point P^* is the common point of all the lines tangent to the absolute at points of S.

We provide an argument, in the $n = 2$ case, to fix ideas. In the double cover, a plane of $\mathbb{H}^2$ corresponds to the intersection of the pseudo-sphere

$$\mathcal{H}^2 = \{x \in \mathbb{R}^3 \mid b_{2,1}(x, x) = -1\}$$

with a time-like plane P of $\mathbb{M}\mathrm{in}^3$, which meets the isotropic cone $\mathcal{I}(b_{2,1})$ of $\mathbb{M}\mathrm{in}^3$ along two light-like vectors v_1, v_2. The light-like planes tangent to the isotropic cone containing v_1, v_2 meet along a line directed by a space-like vector v. It is easy to see that v is orthogonal to v_1 and v_2, and hence to P.

The same holds in anti-de Sitter space: the dual of a space-like plane is a point of $\mathbb{A}\mathrm{d}\mathbb{S}^3$, and vice versa, see Figure 8.

The duality points/hyperplanes suffices to recover the dual of a convex set, see Figure 9. Also, it gives the following description of de Sitter space.

Fact 2.17. *De Sitter space* $\mathrm{d}\mathbb{S}^n$ *is the space of* (*unoriented*) *hyperplanes of* $\mathbb{H}^n$.

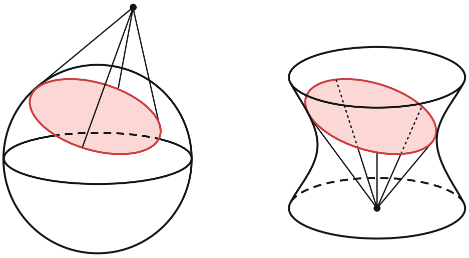


Figure 8. The dual of a plane of the hyperbolic space and of anti-de Sitter space in an affine model

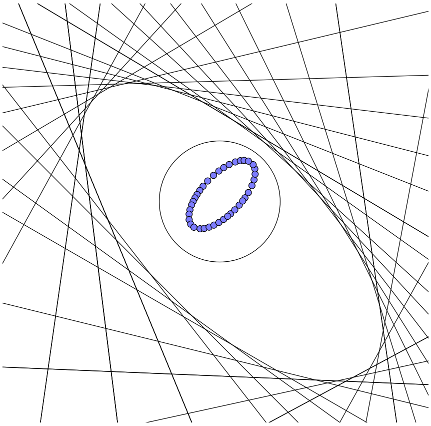


Figure 9. The dual of a convex set with respect to the sphere is the envelope of the hyperplanes dual to its boundary points

Remark 2.18. In the elliptic plane $\mathbb{Ell}^2$, it is well known that the duality can also be expressed in a metric way: the plane x^* is the set of points at distance $\frac{\pi}{2}$ from x. This is readily seen because x^* is the projective quotient of a linear plane P in $\mathbb{E}^3$, and x is the projective quotient of a vector orthogonal to P. The orthogonality for $b_{3,0}$ immediately gives that the projective distance is equal to $\frac{\pi}{2}$ due to (b1) in Section 2.2.

A similar argument leads to the following in $\mathbb{A}\mathrm{d}\mathbb{S}^3$: the plane x^* dual to the point x is the set of points at distance $\frac{\pi}{2}$ from x (see Figure 8).

The same computation also occurs in de Sitter space, if one considers the duality in the following way. As before, let P be a hyperplane in $\mathbb{H}^n$ and let P^* be its dual point in $\mathrm{d}\mathbb{S}^n$. Actually P^* is also dual to a time-like hyperplane of $\mathrm{d}\mathbb{S}^n$, the one defined by the same affine hyperplane as P, and that we still denote by P. Then in $\mathrm{d}\mathbb{S}^n$, the distance between P^* and P is $\frac{\pi}{2}$, see Figure 7.

Comments and references

- When the pseudo-metric is written under the form given by Lemma 2.10, it is usually called a *Hilbert* or *Cayley–Klein* metric. It can be defined on any convex sets, and not only on the ones bounded by quadrics as here, see e.g. [50] for more details.

- As Euclidean space is made only of parabolic lines, it is sometimes called *parabolic space*. In dimension 2, the projective distance can be defined directly on a quadric: this is the Poincaré quadric geometry [52] and [2] (see also the Prologue of the present volume). In particular, this justifies the term “parabolic.”

- Remarkably, in dimension 3, the sphere $\mathcal{S}^3$ has a group structure, namely SU(2), with the determinant as the quadratic form on the ambient space. Similarly, $\mathcal{A}d\mathcal{S}^3$ is $\mathrm{SL}(2,\mathbb{R})$, with minus the determinant as the quadratic form on the ambient space. Note that the Lie algebra $\mathfrak{su}(2)$ (resp. $\mathfrak{sl}(2,\mathbb{R})$) together with its Killing form is naturally identified with $\mathbb{E}^3$ (resp. $\mathbb{Min}^3$).

- The following fact comes readily from the definition of duality: *any line from x meets x^* orthogonally*. This is useful in practice, as the affine models are certainly not conformal (to the Euclidean metric), and computations of angles may be cumbersome, but orthogonality is easily seen, see Figure 7. See Figure 10 for an application. In particular, the Klein model of hyperbolic space is not a conformal model, as a striking difference with the other famous *Poincaré model*, which will not be used in this survey.

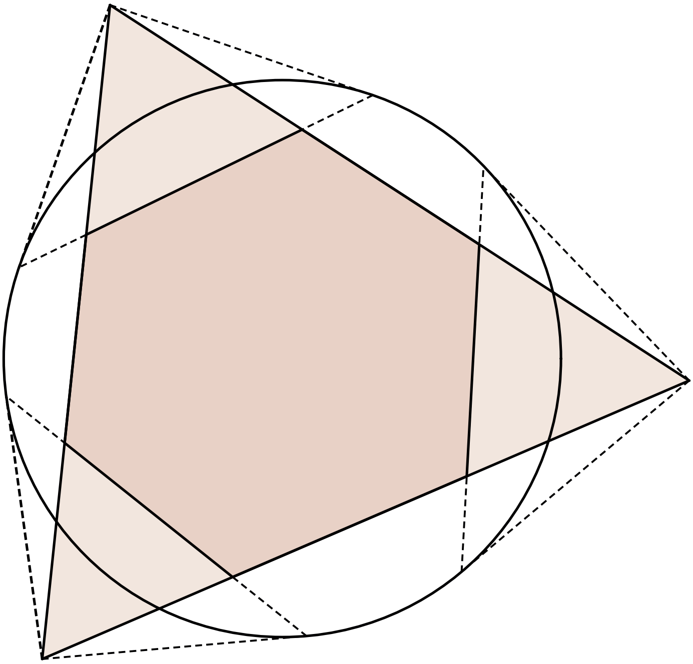


Figure 10. A hyperbolic right-angled hexagon is an affine triangle, hence the corresponding moduli space has dimension 3. Right-angled hexagons are fundamental pieces to construct compact hyperbolic surfaces [66] and [65].

- The duality in $\mathbb{S}^2$ induces also a correspondence between angles and length, that has a great importance as many statements have a straightforward dual analogue. The most basic ones are spherical trigonometric laws for triangles.

This is also true for a general duality with respect to a symmetric non-degenerate bilinear form b. Actually this property is contained in the definition of angle. Let P_1, P_2 be two hyperplanes in $\mathcal{M}$, and let $x \in P_1 \cap P_2$. The outward unit normals n_1 and n_2 to P_1 and P_2 in $T_x\mathcal{M}$ define two points on a pseudo-sphere S of $T_x\mathcal{M}$, which is naturally isometric to a pseudo-Euclidean space. The (exterior dihedral) *angle* between P_1 and P_2 is the pseudo-distance on S between those two points. This does not depend on the choice of the point $x \in P_1 \cap P_2$. But $T_x\mathcal{M}$ is also identified with a hyperplane in the ambient $\mathbb{R}^{n+1}$, the one orthogonal to x for b. In particular, n_1 and n_2 are identified with points in a pseudo-sphere $\mathcal{N}$, dual to P_1 and P_2. The distance between those points of $\mathcal{N}$ is exactly the distance between the points in S, hence equal to the angle between P_1 and P_2.

The point is that in $T_x\mathbb{M}$, n_1 and n_2 may belong to two different pseudo-spheres, but it is possible to define a pseudo-distance between points belonging to two different pseudo-spheres defined by a same symmetric non-degenerate bilinear form. We will not use this in the present paper. For trigonometric laws for hyperbolic/anti-de Sitter triangles, see [65] and [23].

- Let Isom_0 be the connected component of the identity of the isometry group. There are famous identifications between $\mathrm{Isom}_0(\mathbb{H}^2)$ and $\mathrm{PSL}(2,\mathbb{R})$, and between the absolute and $\mathbb{R}\mathrm{P}^1$. In dimension 3, there is an identification between $\mathrm{Isom}_0(\mathbb{H}^3)$ and $\mathrm{PSL}(2,\mathbb{C})$. The last group also acts on the absolute, which is naturally identified with $\mathbb{C}\mathrm{P}^1$. In contrast,

$$\mathrm{Isom}_0(\mathbb{A}\mathrm{d}\mathbb{S}^3) = \mathrm{PSL}(2,\mathbb{R}) \times \mathrm{PSL}(2,\mathbb{R}),$$

and the absolute is naturally identified with $\mathbb{R}\mathrm{P}^1 \times \mathbb{R}\mathrm{P}^1$.

3 Degenerate cases

3.1 Geometries

A *geometry* is a pair (X, G) where X is a manifold and G is a Lie group acting transitively by diffeomorphisms on X. The model spaces we introduced in Section 2 are geometries, with $X = \mathbb{M}$ and $G = \mathrm{PO}(p,q)$. Note that here the group G is a subgroup of the group of projective transformations that can be characterized in two ways: G leaves $\mathbb{M}$ invariant in the projective space and G is the isometry group of $\mathbb{M}$. We will now be interested by *degenerate* model spaces. This means that they are defined by a degenerate bilinear form (or equivalently, their absolute is not a proper quadric in an affine chart).[13]

A degenerate model space gives a geometry $(\mathbb{M}, G)$, with G the subgroup of the group of projective transformations that leaves $\mathbb{M}$ invariant. Equivalently, G is a subgroup of the group of projective transformations that preserves a degenerate $(0,2)$ tensor on $\mathbb{M}$. But we will consider the geometries $(\mathbb{M}, H)$, with H a proper subgroup of G. The choice of H will be justified by a duality argument in this section, and next justified on the one hand by the process of degeneration introduced in Section 4 and on the other hand by the definition of connections and volume forms in Section 5.

Unlike the preceding section, we do not attempt to give a unified treatment of the degenerate geometries, but we focus on Euclidean and Minkowski spaces, together with their dual spaces: co-Euclidean space and co-Minkowski space.

---

[13] In particular, the absolute will contain lines, called *isotropic*.

3.2 Euclidean space

Euclidean space as a degenerate model geometry. Euclidean space $\mathbb{E}^n$ may be considered as a projective quotient of a pseudo-sphere $\mathcal{E}uc^n$ of $\mathbb{R}^{n+1}$ defined by the following degenerate bilinear form:

$$b(x,y) = x_{n+1}y_{n+1}, \tag{7}$$

i.e.

$$\mathcal{E}uc^n = \{x \in \mathbb{R}^{n+1} \mid x_{n+1}^2 = 1\}.$$

Indeed, $\mathcal{E}uc^n/\{\pm \mathrm{Id}\}$ is the complement of a hyperplane in $\mathbb{R}\mathrm{P}^n$, i.e. it is identified with the affine space $\mathbb{A}^n$ of dimension n. Note that in this model, the absolute (the projective quotient of the isotropic cone of b) is the hyperplane at infinity. In particular, all the lines are parabolic, and hence the projective distance is zero, so we do not recover the Euclidean metric. We need more information to recover Euclidean geometry from this model.

The group G of projective transformations leaving $\mathcal{E}uc^n$ invariant is the group of transformations of the form

$$\left(\begin{array}{ccc|c} & & & t_1 \\ & A & & \vdots \\ & & & t_n \\ \hline 0 & \cdots & 0 & \lambda \end{array}\right) \tag{8}$$

where A belongs to $\mathrm{GL}(n,\mathbb{R})$ and $\lambda \neq 0$, quotiented by the subgroup of multiples of the identity, i.e. it is the group of transformations of the form

$$\left[\begin{array}{ccc|c} & & & t_1 \\ & A & & \vdots \\ & & & t_n \\ \hline 0 & \cdots & 0 & \lambda \end{array}\right].$$

Of course one can always choose a representative with $\lambda = 1$. Under the identification of $\mathcal{E}uc^n/\{\pm \mathrm{Id}\}$ with the affine space $\mathbb{A}^n$ given by taking the chart $\{x_{n+1} = 1\}$, the action of G is identified with the action on $\mathbb{A}^n \times \{1\}$ of matrices of the form

$$\left(\begin{array}{ccc|c} & & & t_1 \\ & A & & \vdots \\ & & & t_n \\ \hline 0 & \cdots & 0 & 1 \end{array}\right). \tag{9}$$

Hence $(\mathcal{E}uc^n/\{\pm \mathrm{Id}\}, G)$ is the affine geometry $(\mathbb{A}^n, \mathrm{GL}(n,\mathbb{R}) \rtimes \mathbb{R}^n)$. The Euclidean geometry $(\mathbb{E}^n, \mathrm{Isom}(\mathbb{E}^n))$ is $(\mathcal{E}uc^n/\{\pm \mathrm{Id}\}, H)$, where H is the subgroup of G such that A in (8) belongs to $\mathrm{O}(n)$.

An element of the form (8) also acts on $\mathbb{R}^n \times \{0\}$, and the action reduces to the action of A. As $A \in \mathrm{O}(n)$, passing to the projective quotient, the hyperplane at infinity of $\mathbb{A}^n$ is endowed with the elliptic geometry $\mathbb{Ell}^{n-1}$. Conversely, suppose that $(\mathbb{A}^n, H)$ is a geometry, where H is a group of affine transformations, whose action on the hyperplane at infinity is the one of $\mathrm{PO}(n)$. Then necessarily the part A, in a representative of the form (8) for an element of H, must belong to $\mathrm{O}(n)$. In other words, the Euclidean geometry can be characterized as follows.

Fact 3.1. *Euclidean space is projective space minus an elliptic hyperplane.*

Duality of convex sets. Recall that a *convex body* of $\mathbb{E}^n$ is a compact convex set in $\mathbb{E}^n$ with non-empty interior.[14] We will add the following assumption.

Definition 3.2. A convex body is *admissible* if it contains the origin in its interior.

Actually every convex body is admissible up to a translation. It is suitable to consider K in $\mathbb{E}^n \times \{1\} \subset \mathbb{E}^{n+1}$, in order to introduce the cone $\mathcal{C}(K)$ in $\mathbb{E}^{n+1}$ over K:

$$\mathcal{C}(K) = \left\{ \lambda \begin{pmatrix} x \\ 1 \end{pmatrix} \;\middle|\; x \in K, \lambda \geq 0 \right\}.$$

Let $\mathcal{C}(K)^*$ be its dual in $\mathbb{E}^{n+1}$ for the scalar product $b_{n+1,0}$:

$$\begin{aligned} \mathcal{C}(K)^* = \{(y, y_{n+1}) \in \mathbb{R}^{n+1} \mid b_{n+1,0}((y, y_{n+1}), (x, x_{n+1})) \leq 0, \\ \text{for all } (x, x_{n+1}) \in \mathcal{C}(K)\}. \end{aligned} \tag{10}$$

We will denote by K^* the intersection of $\mathcal{C}(K)^*$ with $\{y_{n+1} = -1\}$. We identify $\{y_{n+1} = -1\}$ with $\mathbb{E}^n$, so that K^* is a closed convex set of $\mathbb{E}^n$. It is readily seen that

$$K^* = \{y \in \mathbb{R}^n \mid b_{n,0}(x, y) \leq 1, \text{ for all } x \in K\}. \tag{11}$$

The expression (11) corresponds to the affine duality with respect to the unit sphere: a point rv, with $v \in \mathcal{S}^n$, on the boundary of K will correspond to a support plane of K^* of direction orthogonal to v and at distance $\frac{1}{r}$ from the origin. Compare Figure 11 and Figure 12. In particular, the dual B_r of a ball centred at the origin with radius r is the ball $B_{1/r}$ centred at the origin.

---

[14] For some authors, a convex body is only a compact convex set.

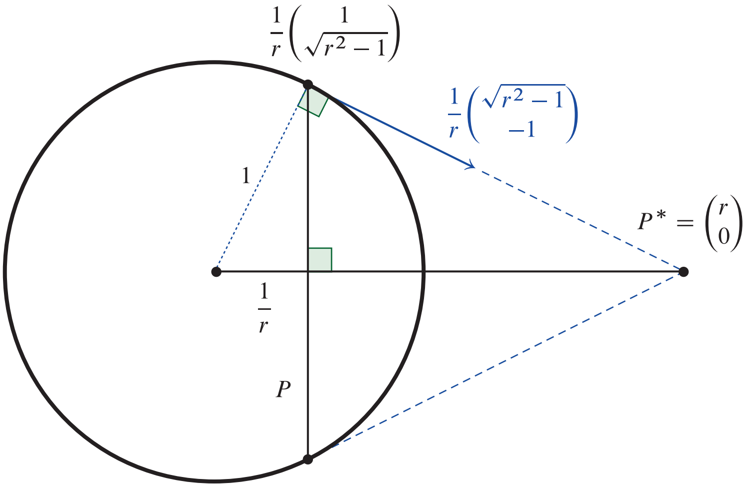

$\frac{1}{r}\begin{pmatrix}1\\ \sqrt{r^2-1}\end{pmatrix}$
$\frac{1}{r}\begin{pmatrix}\sqrt{r^2-1}\\ -1\end{pmatrix}$
1
$P^* = \begin{pmatrix}r\\ 0\end{pmatrix}$
$\frac{1}{r}$
P

Figure 11. The Euclidean duality is the duality with respect to the unit sphere

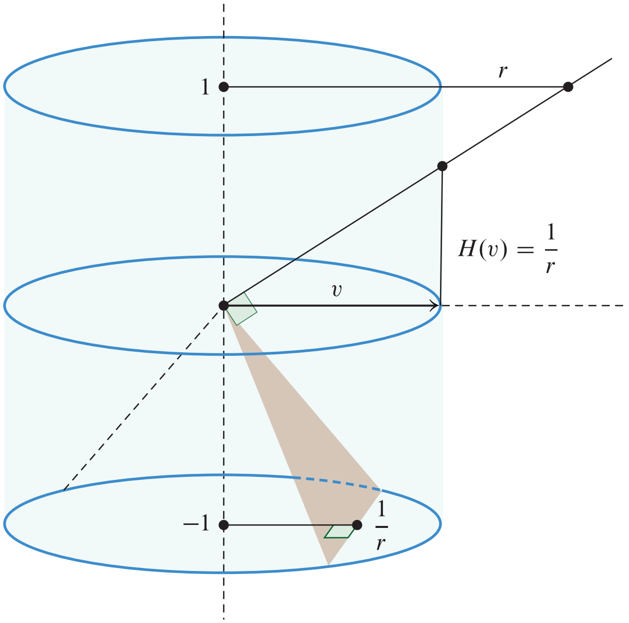


Figure 12. The duality between points and lines in the Euclidean plane expressed in terms of orthogonality in a higher-dimensional Euclidean space

Lemma 3.3. *The dual of an admissible convex body in Euclidean space is an admissible convex body in Euclidean space.*

Proof. If K is a convex body, by definition there exists $s, t > 0$ such that one has $B_s \subset K \subset B_t$. Hence by (4), $B_{1/t} \subset K^* \subset B_{1/s}$: K^* is bounded and contains the origin in its interior. □

Support functions. Let $H\colon \mathbb{R}^n \to \mathbb{R}$ be such that $\mathcal{C}(K)$ is the epigraph of H, i.e.

$$\mathcal{C}(K) = \{(x, x_{n+1}) \in \mathbb{R}^{n+1} \mid x_{n+1} \geq H(x)\}. \tag{12}$$

The function H is convex, positive outside the origin, and homogeneous of degree 1: $H(\lambda x) = \lambda H(x)$ for $\lambda > 0$. Hence using (12) and (10),

$$\begin{aligned}\mathcal{C}(K)^* = \{(y, y_{n+1}) \in \mathbb{R}^{n+1} \mid b_{n,0}(y, x) \leq -y_{n+1}x_{n+1},\\ \text{for all } (x, x_{n+1}), x_{n+1} \geq H(x)\}\end{aligned}$$

i.e.

$$\mathcal{C}(K)^* = \{(y, y_{n+1}) \in \mathbb{R}^{n+1} \mid b_{n,0}(y, x) \leq -y_{n+1}H(x), \text{ for all } x \in \mathbb{R}^n\}$$

so K^* is determined by H:

$$K^* = \{y \in \mathbb{R}^n \mid b_{n,0}(y, x) \leq H(x), \text{ for all } x \in \mathbb{R}^n\}. \tag{13}$$

Definition 3.4. The function H such that $\mathcal{C}(K)$ is the epigraph of H is the *support function* of K^*.

The support function has also the following interpretation. Let v be a unit vector of $\mathbb{E}^n \times \{0\}$. Then $H(v)$ is the distance in $\mathbb{E}^n \times \{-1\}$ between $\{0\} \times \{-1\}$ and the support plane of K^* directed by v (see Figure 12). Hence (13) expresses the fact that K^* is the envelope of its support planes.

The following fact follows easily from the construction.

Fact 3.5. *The support function provides a bijection between the space of admissible convex bodies of $\mathbb{R}^n$ and the space of positive convex 1-homogeneous functions on $\mathbb{R}^n$.*

3.3 Co-Euclidean space

Definition. On $\mathbb{R}^{n+1}$, let b^* be the following degenerate bilinear form

$$b^*(x, y) = x_1 y_1 + \cdots + x_n y_n.$$

Let $co\mathcal{E}uc^n$ be the unit sphere for b^*:

$$co\mathcal{E}uc^n = \{x \in \mathbb{R}^{n+1} \mid b^*(x, x) = 1\},$$

endowed with the restriction of b^* to its tangent space. Topologically, $co\mathcal{E}uc^n$ is $\mathcal{S}^n \times \mathbb{R}$.

Definition 3.6. The space ${}^*\mathbb{E}^n = co\mathcal{E}uc^n/\{\pm \mathrm{Id}\}$ is *co-Euclidean space*.

The isotropic cone $\mathcal{I}(b^*)$ is the line $\tilde{\ell} = \{(0, \ldots, 0, \lambda)\}$. Let $\ell = \mathrm{P}\tilde{\ell}$. Then ${}^*\mathbb{E}^n = \mathbb{R}\mathrm{P}^n \setminus \{\ell\}$. Lines of ${}^*\mathbb{E}^n$ are parabolic if they contain ℓ, and elliptic otherwise.

Duality. Co-Euclidean space is dual to the Euclidean space in the sense that it can be described as the set of hyperplanes of $\mathbb{E}^n$. Let P be an affine hyperplane in $\mathbb{E}^n$, and v its unit normal vector, pointing towards the side of P which does not contain the origin 0 of $\mathbb{E}^n$. Let $h(v)$ be the distance from 0 to P, i.e. P has equation $b_{n,0}(\cdot, v) - h(v) = 0$. The vector $\binom{v}{h(v)} \in \mathbb{R}^{n+1}$ is orthogonal in $\mathbb{E}^{n+1}$ to the linear hyperplane containing $P \times \{-1\}$. Its projective quotient defines a point P^* in ${}^*\mathbb{E}^n$, see Figure 12. One could also consider the other unit normal vector $-v$ of P. The (signed) distance from the origin is then $-h(v)$, and the point $-\binom{v}{h(v)}$ has the same projective quotient as $\binom{v}{h(v)}$.

Fact 3.7. *Co-Euclidean space ${}^*\mathbb{E}^n$ is the space of (unoriented) hyperplanes of $\mathbb{E}^n$.*

Conversely, an elliptic hyperplane P of ${}^*\mathbb{E}^n$ (i.e. a hyperplane which does not contain ℓ) is dual to a point P^* of $\mathbb{E}^n$. The point P^* is the intersection of all the hyperplanes x^*, for $x \in P$. A co-Euclidean hyperplane of ${}^*\mathbb{E}^n$ (i.e. a hyperplane containing ℓ) is dual to a point at infinity.

Let K be an admissible convex body of $\mathbb{E}^n$, and let H be its support function, i.e. $\mathcal{C}(K)^*$ is the epigraph of H. By abuse of notation, let us also denote by K^* the intersection of $\mathcal{C}(K)^*$ with $co\mathcal{E}uc^n$. The set K^* is the epigraph of the restriction h of H to $co\mathcal{E}uc^n \cap \{x_{n+1} = 0\}$, that we identify with $\mathcal{S}^{n-1}$. Note that by homogeneity, H is determined by its restriction to $\mathcal{S}^{n-1}$, which is actually h:

$$H(x) = \|x\| h\Big(\frac{x}{\|x\|}\Big),$$

with $\|x\| = \sqrt{b_{n,0}(x, x)}$.

Fact 3.8. *The convex set K^* of co$\mathcal{E}uc^n$ is the epigraph of h, the restriction to $\mathcal{S}^{n-1}$ of the support function of K.*

Remark 3.9. The duality between hyperplanes of $\mathbb{E}^n$ and points in ${}^*\mathbb{E}^n$ leads to the following relation between angles and length. If P' is another affine hyperplane of $\mathbb{E}^n$, orthogonal to the unit vector v', then the ${}^*\mathbb{E}^n$ segment between P^* and P'^* is elliptic (or equivalently, the restriction of b^* is positive definite on any non-vertical hyperplane of $\mathbb{R}^{n+1}$). By the expression (b1) in Section 2.2, $\cos d(P^*, P'^*) = b^*(P^*, P'^*)$ and it is readily seen that the last quantity is equal to $b_{n,0}(v, v')$. Thus, the (exterior dihedral) angles between intersecting affine hyperplanes of $\mathbb{E}^n$ are equal to the distance between their duals in ${}^*\mathbb{E}^n$.

A model geometry. Let $\mathrm{Isom}(b^*)$ be the subgroup of projective transformations that preserve b^*. There is a natural injective homomorphism

$$^*\colon \mathrm{Isom}(\mathbb{E}^n) \longrightarrow \mathrm{Isom}(b^*)$$

which is defined as follows. Let

$$P(v,h) = \{z \in \mathbb{R}^n \mid b_{n,0}(z,v) = h\}$$

be an affine hyperplane of $\mathbb{E}^n$, $v \in \mathcal{S}^{n-1}, h \in \mathbb{R}^*$. For $A \in O(n)$ and $\vec{t} \in \mathbb{R}^n$, we have

$$AP(v,h) + \vec{t} = P(Av, h + b_{n,0}(v, A^{-1}\vec{t})),$$

from which we define

$$^*\left[\begin{array}{c|c} \Large A & \vec{t} \\ \hline 0 \ \cdots \ 0 & 1 \end{array}\right] := \left[\begin{array}{c|c} \Large A & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \\ \hline A^{-1}\vec{t} & 1 \end{array}\right]. \tag{14}$$

Definition 3.10. The isometry group of co-Euclidean space, $\mathrm{Isom}({}^*\mathbb{E}^n)$, is the group of projective transformations of the form

$$\left[\begin{array}{c|c} \Large A & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \\ \hline \vec{t} & 1 \end{array}\right] \tag{15}$$

for $A \in \mathrm{O}(n), \vec{t} \in \mathbb{R}^n$.

Note that Isom($^*\mathbb{E}^n$) is a proper subgroup of the group of isometries of b^*. For instance, the latter also contains *homotheties* of the form $[\mathrm{diag}(1, \dots, 1, \lambda)]$. They correspond to displacement along parabolic lines, see Figure 13.

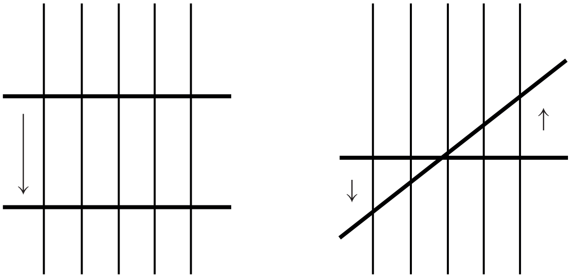


Figure 13. In this affine model of the co-Euclidean plane, the absolute $\{\ell\}$ is at infinity, in such a way that the vertical lines are parabolic. Also the quotient of $\{1, -1\} \times \mathbb{R} \subset \mathbb{S}^1 \times \mathbb{R}$ is at infinity. On the left is the effect of the dual of a Euclidean homothety on an elliptic line of the co-Euclidean plane, and on the right is the effect of the dual of a Euclidean translation on an elliptic line of the co-Euclidean plane.

Co-Euclidean space is naturally endowed with a degenerate metric g^*, which is the restriction on $co\mathcal{E}uc^n$ of b^*, pushed down to the projective quotient, and coincides with the elliptic metric on the elliptic hyperplanes, and is zero on the parabolic lines.

The isotropy group $I(p)$ of a point $p \in {}^*\mathbb{E}^n$ corresponds to the isometries of $\mathbb{E}^n$ that fix the dual hyperplane p^*, i.e. all the translations by vectors in the direction of p^*. The group $I(p)$ actually pointwise fixes the parabolic line passing through p, because each point on the line is dual to a hyperplane parallel to p^* in $\mathbb{E}^n$. By a similar argument, $I(p)$ preserves every other parabolic line (not through p) and acts on it by translation. Finally, the group $I(p)$ acts simply transitively on all the elliptic hyperplanes containing p, because two such hyperplanes a, b are dual to two points a^*, b^* contained in the hyperplane p^* of $\mathbb{E}^n$, and there is a unique translation in the direction of p^* sending a^* to b^*. The stabilizer in $I(p)$ of any such elliptic hyperplane clearly acts on it as $\mathrm{O}(n-1)$.

As a result of this discussion, let us show that any $(0, 2)$-tensor on $^*\mathbb{E}^n$ invariant under the action of Isom($^*\mathbb{E}^n$) must necessarily be degenerate. Let g be a bilinear form on $T_p{}^*\mathbb{E}^n$ invariant under the action of $I(p)$, and fix a vector $X \in T_p{}^*\mathbb{E}^n$ which is not tangent to the parabolic line through p. By the above discussion, there is an element of $I(p)$ which maps X to $X + V$, where V is a vector tangent to the parabolic line. Hence if $I(p)$ preserves the bilinear form g, then V is null for g. Hence g cannot be a scalar product. Since $I(p)$ acts transitively on lines spanned by vectors X as above, all such vectors X must be of the same type, and therefore g is degenerate.

Fact 3.11. *There is no pseudo-Riemannian metric on* $({}^*\mathbb{E}^n, \mathrm{Isom}({}^*\mathbb{E}^n))$.

In fact, the argument above essentially shows that the metric must be of the form described above (up to a factor), namely when lifted to the double cover $co\mathcal{E}uc^n$, it restricts to the spherical metric on every hyperplane transverse to the parabolic line.

3.4 Minkowski space

Minkowski space as a degenerate model geometry. Like Euclidean space, Minkowski space is a subgeometry of the affine geometry, and isometries of Minkowski space are of the form

$$\left[\begin{array}{ccc|c} & & & t_1 \\ & A & & \vdots \\ & & & t_n \\ \hline 0 & \cdots & 0 & 1 \end{array}\right],$$

with $A \in \mathrm{O}(n-1,1)$. Such a transformation also acts on $\mathbb{R}^n \times \{0\}$, and the action reduces to the action of A. As $A \in \mathrm{O}(n-1,1)$, passing to the projective quotient, the hyperplane at infinity of $\mathbb{A}^n$ is endowed with the hyperbolic and the de Sitter geometries. Conversely, suppose that $(\mathbb{A}^n, H)$ is a geometry, where H is a group of affine transformations, whose action on the hyperplane at infinity is the one of $\mathrm{PO}(n-1,1)$. Then necessarily the part A for the representative of an element of H must belong to $\mathrm{O}(n-1,1)$. This can be summarized as follows.

Fact 3.12. *Minkowski space is the projective space minus a hyperbolic-de Sitter hyperplane.*

Duality of convex sets. An affine space-like hyperplane P splits $\mathbb{M}\mathrm{in}^n$ into two half-spaces. The time-orientation of $\mathbb{M}\mathrm{in}^n$ allows to speak about the future side of P. A convex set is *future convex* if it is the intersection of the future of space-like hyperplanes. Note that a future convex set may have also light-like support planes (e.g. the future cone of a point),[15] but no time-like support plane.

Fact 3.13. *A future convex set contains the future cone of any of its points.*

---

[15] The *future cone of a point* is the union of all the future directed time-like or light-like half lines from the point.

Let us denote by $\mathcal{F}$ the closure of the future cone of the origin minus the origin, i.e.

$$\mathcal{F} := \{x \in \mathbb{R}^n \mid b_{n-1,1}(x,x) \leq 0, x_n > 0\}. \tag{16}$$

Definition 3.14. An *admissible convex set* of $\mathbb{Min}^n$ is a future convex set contained in $\mathcal{F}$.

Up to translation, any future convex set contained in the future cone of a point is an admissible convex set. But not all future convex sets are admissible, even up to translation, for example consider the future of a single space-like hyperplane.

Let K be an admissible convex set of $\mathbb{Min}^n$, and let us identify $\mathbb{Min}^n$ with $\{-1\}\times\mathbb{Min}^n$ in $\mathbb{Min}^{n+1}$. Then the cone $\mathcal{C}(K)$ in $\mathbb{Min}^{n+1}$ over K is

$$\mathcal{C}(K) = \left\{\lambda\begin{pmatrix}-1\\ x\end{pmatrix} \,\middle|\, x \in K, \lambda \geq 0\right\}.$$

Note that $C(K)$ is not closed as it contains points with zero first coordinate in its closure.

Let $\mathcal{C}(K)^*$ be its dual in $\mathbb{Min}^{n+1}$ for $b_{n,1}$:

$$\mathcal{C}(K)^* = \{(y_1,y) \in \mathbb{R}^{n+1} \mid b_{n,1}((y_1,y),(x_1,x)) \leq 0, \text{ for all } (x_1,x) \in \mathcal{C}(K)\}. \tag{17}$$

We will denote by K^* the intersection of $\mathcal{C}(K)^*$ with $\{y_1 = -1\}$. As we identify $\{y_1 = -1\}$ with $\mathbb{Min}^n$, K^* is a closed convex set of $\mathbb{Min}^n$, and it is readily seen that

$$K^* = \{y \in \mathbb{R}^n \mid b_{n-1,1}(x,y) \leq -1, \text{ for all } x \in K\}. \tag{18}$$

This corresponds to the affine duality with respect to the unit hyperboloid, compare Figure 14 and Figure 15. The dual of H_r, the future convex side of a branch of hyperboloid with radius r, is $H_{1/r}$. The fact that $H_1^* = H_1$ comes also from the following: the boundary of $\mathcal{C}(H_1)$ is (a part of) the isotropic cone $\mathcal{I}(b_{n,1})$, which is dual to itself for $b_{n,1}$.

Let us call *admissible cone* the future cone of a point contained in the interior of $\mathcal{F}$, and *admissible truncation* the intersection of $\mathcal{F}$ with the future side of a space-like hyperplane P such that the origin is in the past of P, see Figure 16.

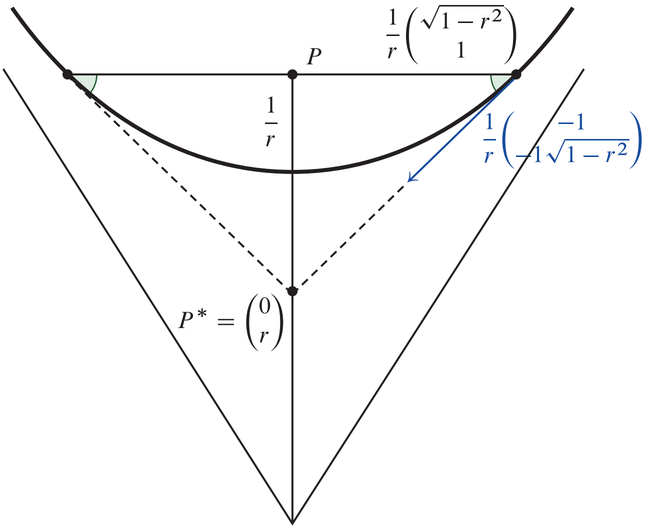

$\frac{1}{r}\begin{pmatrix}\sqrt{1-r^2}\\ 1\end{pmatrix}$
P
$\frac{1}{r}$
$\frac{1}{r}\begin{pmatrix}-1\\ -1\sqrt{1-r^2}\end{pmatrix}$
$P^* = \begin{pmatrix}0\\ r\end{pmatrix}$

Figure 14. The duality of admissible convex sets in Minkowski space is the affine duality with respect to the upper hyperboloid

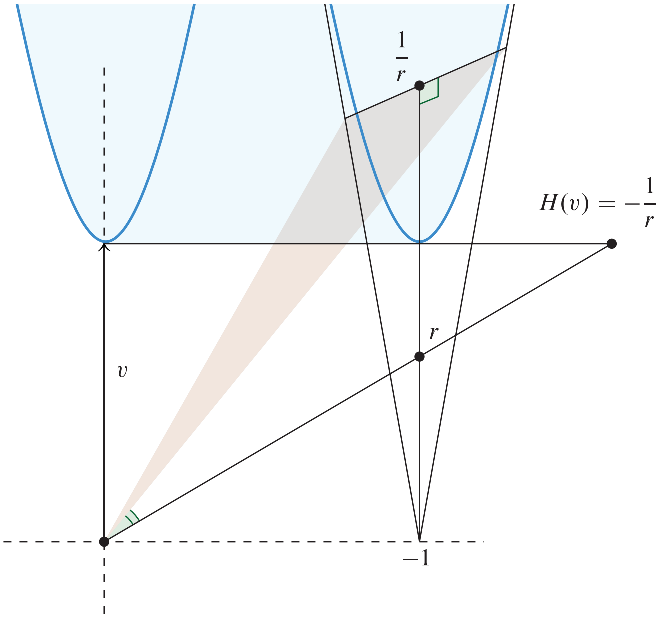


Figure 15. The duality between points and lines in the Minkowski plane expressed in term of orthogonality in a higher-dimensional Minkowski space

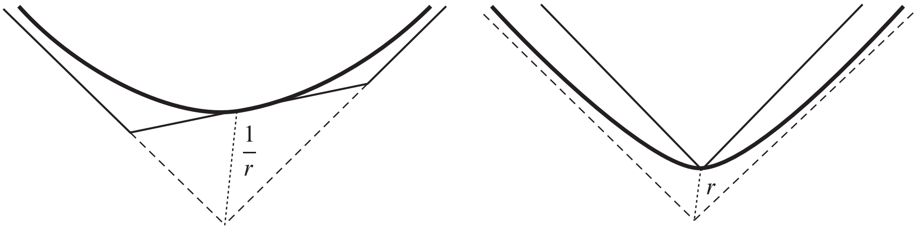


Figure 16. The dual of an admissible truncation is an admissible cone

Lemma 3.15. *The dual of an admissible truncation is an admissible cone, and vice versa.*

Proof. Let K be an admissible truncation, i.e. if v is the unit future vector orthogonal to P and if r is the distance from the origin to P, then

$$K = \{x \in \mathbb{Min}^n \mid b_{n-1,1}(x,x) < 0, b_{n-1,1}(x - rv, v) \leq 0\}.$$

From (18), we see that $\frac{1}{r}v \in K^*$. From Fact 3.13, if C is the future cone of $\frac{1}{r}v$, then $C \subset K^*$.

On the other hand, as $\frac{1}{r}v \in C$, if $y \in C^*$, then by (18), $b_{n-1,1}(y,v) \leq -r$, which implies that $y \in K$, hence $C^* \subset K$. The result follows from (6) and (4). ☐

Lemma 3.16. *The dual of an admissible convex set of Minkowski space is an admissible convex set of Minkowski space.*

Proof. Let K be an admissible convex set. By Fact 3.13, K contains the future cone of a point, and by (4) and Lemma 3.15, K^* is contained in an admissible truncation, in particular K^* is in $\mathcal{F}$.

As K is in $\mathcal{F}$, there exists a hyperplane separating K from the origin, in particular K is contained in an admissible truncation. By (4) and Lemma 3.15, K^* contains the future cone of a point. From this, it is easy to see that the closed convex set K^* must be a future convex set. ☐

Support function. Let $\widetilde{H}\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be the convex function whose graph is the cone over the boundary of K. In particular, $\mathcal{C}(K)$ is the intersection of the epigraph of $\widetilde{H}$ with $\{x_1 > 0\}$:

$$\mathcal{C}(K) = \{(x_1, x) \in \mathbb{R}_{>0} \times \mathbb{R}^n \mid x_1 \geq \widetilde{H}(x)\}. \tag{19}$$

The function $\widetilde{H}$ is non-positive, convex and homogeneous of degree one. Let dom $\widetilde{H}$ be the domain of $\widetilde{H}$, the set of points where $\widetilde{H}$ takes finite values. By Fact 3.13, it is easy to see that any future time-like ray from the origin meets the boundary of K exactly once, hence

$$\operatorname{int}\mathcal{F} \subset \operatorname{dom}\widetilde{H} \subset \mathcal{F},$$

(here $\mathcal{F}$ is the set (16) for the Minkowski structure induced on $\{0\} \times \mathbb{R}^n$ by that of the ambient $\mathbb{Min}^{n+1}$). Let $v \in \partial\mathcal{F}$, i.e. v is a future light-like vector. If there exists a $\lambda > 0$ such that $\lambda v \in K$, then $v \in \operatorname{dom}\widetilde{H}$. Otherwise, as from (19),

$$K = \{x \in \mathcal{F} \mid \widetilde{H}(x) \leq -1\}, \tag{20}$$

we would have $\widetilde{H}(\lambda v) > -1$. So, by homogeneity, $0 \geq \widetilde{H}(v) \geq -1/\lambda$ for all $\lambda > 0$, hence $\widetilde{H}(v) = 0$, and the domain of $\widetilde{H}$ is $\mathcal{F}$.

As the epigraph of $\widetilde{H}$ is closed, $\widetilde{H}$ is lower semi-continuous, hence it is determined by its restriction to $\operatorname{int}\mathcal{F}$. See Section 7 in [54] for details. We will denote by H the restriction of $\widetilde{H}$ to $\operatorname{int}\mathcal{F}$. The function H is negative.

Definition 3.17. The function H is the *support function* of K^*.

The exact relation between H and K^* is given by the following lemma.

Lemma 3.18. *Let H be the support function of K^*. Then*

$$K^* = \{y \in \mathbb{R}^n \mid b_{n-1,1}(y, x) \leq H(x), \textit{ for all } x \in \operatorname{int}\mathcal{F}\}. \tag{21}$$

Proof. By (18) and (20)

$$K^* = \{y \in \mathbb{R}^n \mid b_{n-1,1}(x, y) \leq -1 \text{ for all } x \in \mathcal{F}, \widetilde{H}(x) \leq -1\},$$

so it is straightforward that the set

$$A = \{y \in \mathbb{R}^n \mid b_{n-1,1}(x, y) \leq \widetilde{H}(x) \text{ for all } x \in \mathcal{F}\}$$

is contained in K^*. Conversely, suppose that $y \in K^*$ and let $x \in \mathcal{F}$. If $\widetilde{H}(x) = 0$, then $y \in A$. Otherwise, by homogeneity,

$$\widetilde{H}\Big(\frac{x}{-\widetilde{H}(x)}\Big) = -1$$

so

$$b_{n-1,1}\Big(\frac{x}{-\widetilde{H}(x)}, y\Big) \leq -1$$

i.e. $b_{n-1,1}(x, y) \leq \widetilde{H}(x)$, so $y \in A$. By lower-semi continuity, the right-hand side of (21) is equal to A. □

The support function also has the following interpretation. Let v be a unit vector of $\{0\} \times \operatorname{int} \mathcal{F}$. Then $H(v)$ is the distance in $\{-1\} \times \mathbb{M}\mathrm{in}^n$ between the origin and the space-like support plane of K^* directed by v (see Figure 15).[16] Hence (21) expresses the fact that K^* is the envelope of its space-like support planes.

The following fact follows easily from the construction,

Fact 3.19. *The support function provides a bijection between the space of admissible convex subsets of* $\mathbb{M}\mathrm{in}^n$ *and the space of negative convex* 1*-homogenous functions on the cone* $\operatorname{int} \mathcal{F}$.

3.5 Co-Minkowski space

Definition. On $\mathbb{R}^{n+1}$, let b^*_- be the following degenerate bilinear form

$$b^*_-(x, y) = x_2 y_2 + \cdots + x_n y_n - x_{n+1} y_{n+1}.$$

Let $co\mathcal{M}in^n$ be the unit sphere for b^*_-:

$$co\mathcal{M}in^n = \{x \mid b^*_-(x, x) = -1\},$$

endowed with the restriction of b^*_- on its tangent space. Topologically, $co\mathcal{M}in^n$ is the real line times an open disc. Any section of $co\mathcal{M}in^n$ by a hyperplane not containing a horizontal line is isometric to $\mathcal{H}^n$.

Definition 3.20. The space ${}^*\mathbb{M}\mathrm{in}^n = co\mathcal{M}in^n / \{\pm \mathrm{Id}\}$ is *co-Minkowski space*.

The isotropic cone $\mathcal{I}(b^*)$ contains the line $\tilde{\ell} = \{(\lambda, 0, \ldots, 0)\}$. Let us call $\ell = \mathrm{P}\,\tilde{\ell}$ the *vertex* of the absolute. Hyperplanes of co-Minkowski space which do not contain the vertex of the absolute are hyperbolic. Lines passing through the vertex of the absolute are parabolic.

Fact 3.21. *A strip between two parallel lines or the interior of a cone in the plane with the projective distance is a model of* ${}^*\mathbb{M}\mathrm{in}^2$ *or* $\overline{{}^*\mathbb{M}\mathrm{in}^2}$.

The two lines in the fact above meet at the vertex of the absolute. In an affine chart, in dimension 3, we have the following, see Figure 17.

Fact 3.22. *The convex side of a ruled quadric in* $\mathbb{R}^3$ *(i.e. elliptic cone or elliptic cylinder or hyperbolic cylinder) with the projective distance is a model of* ${}^*\mathbb{M}\mathrm{in}^3$ *or* $\overline{{}^*\mathbb{M}\mathrm{in}^3}$.

---

[16] In other terms, the Lorentzian orthogonal projection of the origin onto the support plane is the point $-H(v)v$.

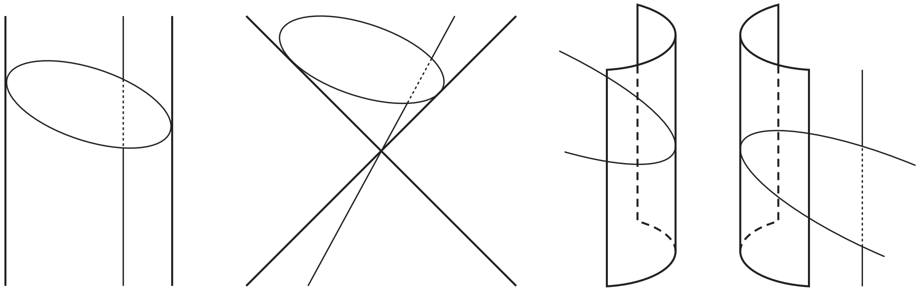


Figure 17. Projectively equivalent affine models of 3d co-Minkowski space. In each case, a hyperbolic plane and a parabolic line are drawn.

Duality. Let P be an affine space-like hyperplane in $\mathbb{Min}^n$, and v its future unit normal vector. Let $h(v)$ be the Lorentzian distance from 0 to P, i.e. the future of P has equation $b_{n-1,1}(\cdot, v) < -h(v)$. The vector $\binom{h(v)}{v} \in \mathbb{R}^{n+1}$ is orthogonal in $\mathbb{Min}^{n+1}$ to the linear hyperplane containing $P \times \{-1\}$. Its projective quotient defines a point P^* in $^*\mathbb{Min}^n$, see Figure 15. One could also consider the past unit normal vector $-v$ of P. The (signed) Lorentzian distance from the origin is then $-h(v)$, and the point $-\binom{h(v)}{v}$ has the same projective quotient as $\binom{h(v)}{v}$.

Fact 3.23. *Co-Minkowski space $^*\mathbb{Min}^n$ is the space of (unoriented) space-like hyperplanes of $\mathbb{Min}^n$.*

Conversely, a hyperbolic hyperplane of $^*\mathbb{Min}^n$ (i.e. a hyperplane of $^*\mathbb{Min}^n$ which does not contain ℓ) is dual to a point of $\mathbb{Min}^n$. A co-Minkowski hyperplane of $^*\mathbb{Min}^n$ (i.e. a hyperplane of $^*\mathbb{Min}^n$ which contains ℓ) is dual to a point at infinity.

Let K be an admissible convex subset of $\mathbb{Min}^n$, and let H be its support function. By abuse of notation, let us also denote by K^* the intersection of $\mathcal{C}(K)^*$ with $co\mathcal{M}in^n$. The convex set K^* is the closure of the epigraph of the restriction h of H to $co\mathcal{M}in^n \cap \{x_1 = 0\}$, which we identify with the upper part of the hyperboloid:

$$\mathcal{H}^{n-1}_+ = \mathcal{H}^{n-1} \cap \{x_n > 0\}.$$

Note that by homogeneity, H is determined by its restriction h to $\mathcal{H}^{n-1}_+$:

$$H(x) = \|x\|_- h(x/\|x\|_-),$$

with $\|x\|_- = \sqrt{-b_{n-1,1}(x,x)}$.

Fact 3.24. *The convex set K^* of $co\mathcal{M}in^n$ is the epigraph of h, the restriction to $\mathcal{H}^{n-1}_+$ of the support function of K.*

Note that the function $h\colon \mathcal{H}^{n-1}_{+} \to \mathbb{R}$ can be extended by symmetry to $\mathcal{H}^{n-1}$, in such a way that the projective quotient is well defined on $\mathbb{H}^{n-1}$.

Cylinder model. There is another convenient way to describe the dual of admissible convex sets of Minkowski space in co-Minkowski space, by looking at the affine chart $\{x_{n+1} = 1\}$. In this model, ${}^*\mathbb{M}\mathrm{in}^n$ is an elliptic cylinder $B^{n-1} \times \mathbb{R}$, with B^{n-1} the open unit disc.

Let H be the support function of K^*. The homogeneous function H is also uniquely determined by its restriction $\bar{h}$ to $B^{n-1} \times \{1\}$. The intersection of $\mathcal{C}(K)$ with $\{x_{n+1} = 1\}$ is the graph of $\bar{h}$. As a restriction of a convex function to a hyperplane, $\bar{h}$ is a convex function. Conversely, it can be easily seen that the 1-homogeneous extension of a convex function $B^{n-1} \times \{1\}$ is a convex function on the future cone of the origin. See e.g. [14, Lemma 2.6].

Fact 3.25. *There is a bijection between negative convex functions $B^{n-1} \to \mathbb{R}$ and admissible convex subsets of Minkowski space $\mathbb{M}\mathrm{in}^n$.*

Let us denote by $K^\#$ the intersection of $C(K)^*$ with $\{x_{n+1} = 1\}$. By (21), the convex function $\bar{h}$ determines $K^\#$:

$$K^\# = \left\{ y \;\middle|\; b_{n-1,1}\left(\begin{pmatrix} x \\ 1 \end{pmatrix}, \begin{pmatrix} y \\ y_n \end{pmatrix}\right) \le \bar{h}(x) \right\}.$$

The way to go from K^* to $K^\#$ is by a projective transformation sending the origin to infinity, see Figure 18.

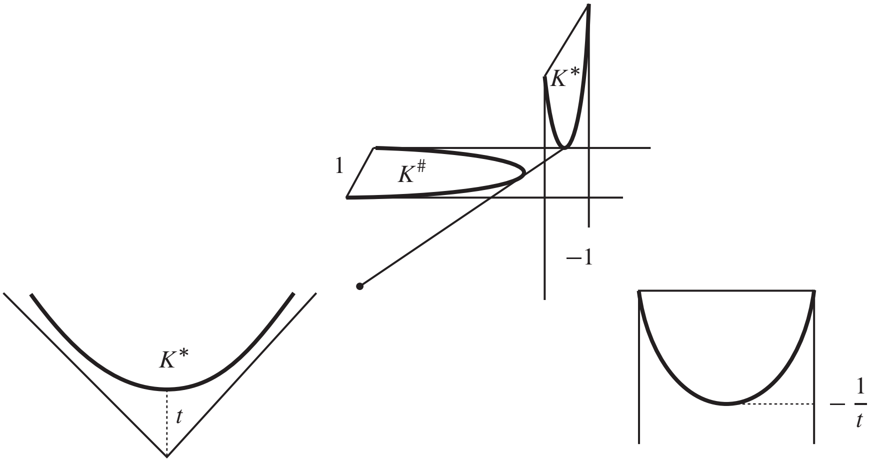


Figure 18. The projective transformation $\binom{x}{y} \mapsto \binom{x/y}{-1/y}$ gives $K^\#$ from K^*. A hyperbola is sent to a parabola.

A model geometry. Let $\mathrm{Isom}(b^*_-)$ be the subgroup of projective transformations that preserves b^*. There is a natural injective homomorphism $^*\colon \mathrm{Isom}(\mathbb{M}\mathrm{in}^n) \to \mathrm{Isom}(b^*_-)$ which is defined as follows. Let

$$P(v,h) = \{z \in \mathbb{R}^n \mid b_{n-1,1}(z,v) = h\}$$

be a space-like affine hyperplane of $\mathbb{M}\mathrm{in}^n$, $v \in \mathcal{H}^{n-1}_+$, $h \in \mathbb{R}^*$. For $A \in \mathrm{O}(n-1,1)$ and $\vec{t} \in \mathbb{R}^n$, we have

$$AP(v,h) + \vec{t} = P(Av, h + b_{n-1,1}(v, A^{-1}\vec{t})),$$

from which we define $^*I$ for $I \in \mathrm{Isom}(\mathbb{M}\mathrm{in}^n)$ as in (14), with $A \in \mathrm{O}(n-1,1)$.

Definition 3.26. The isometry group of co-Minkowski space, $\mathrm{Isom}(^*\mathbb{M}\mathrm{in}^n)$, is the group of projective transformations of the form (15) for $A \in \mathrm{O}(n-1,1)$, $\vec{t} \in \mathbb{R}^n$.

Note that $\mathrm{Isom}(^*\mathbb{M}\mathrm{in}^n)$ is a proper subgroup of the group of isometries of b^*. For instance the latter also contains *homotheties* of the form $[\mathrm{diag}(1,\dots,1,\lambda)]$. As in the co-Euclidean case, these homotheties correspond to displacement along parabolic lines, see Figure 19.

Co-Minkowski space is naturally endowed with a degenerate metric g^*, which is the hyperbolic metric on the hyperbolic hyperplanes, and zero on the parabolic lines.

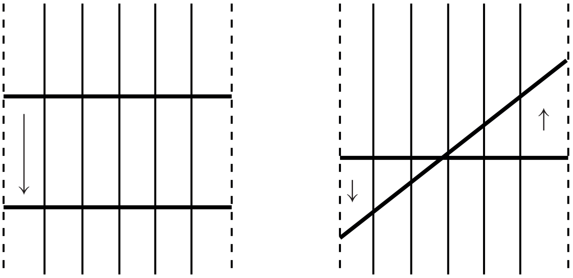


Figure 19. In this affine model of the co-Minkowski plane, the absolute is composed of two vertical lines. The other vertical lines are parabolic. On the left is the effect of a Minkowski homothety on a hyperbolic line of the co-Minkowski plane, and on the right is the effect of a Minkowski translation on a hyperbolic line of the co-Minkowski plane.

Following the same reasoning as in the co-Euclidean case, we see that any $(0,2)$-tensor field on the co-Minkowski space which is invariant under the action of $\mathrm{Isom}(^*\mathbb{M}\mathrm{in}^n)$ must have parabolic directions in its kernel.

Fact 3.27. *There is no pseudo-Riemannian metric on* $(^*\mathbb{M}\mathrm{in}^n, \mathrm{Isom}(^*\mathbb{M}\mathrm{in}^n))$.

Comments and references

- One fundamental property of the support functions is that they behave well under addition. More precisely, the Minkowski sum $A + B = \{a + b \mid A \in A, b \in B\}$ of two admissible convex bodies is an admissible convex body. If H_A and H_B are the corresponding support functions, then one has $H_{A+B} = H_A + H_B$. Classical references for convex bodies are for example [14] and the up-to-date [59].

- From Remark 3.9 and (2), the angle between two lines in the Euclidean plane can be written as the logarithm of a cross-ratio. This is the *Laguerre formula*, see e.g. [53].

- Similarly to Fact 3.1 and 3.12, on has that $(\mathbb{R}^n, b_{n-2,2})$ can be identified with the projective space of dimension n minus a hyperplane endowed with the anti-de Sitter geometry.

- The duality between Euclidean (resp. Minkowski) and co-Euclidean (resp. co-Minkowski) spaces is expressed using a different formalism in [24] and [64].

- Similarly to Fact 3.23, it is easy to see that the outside of co-Minkowski space (i.e. the projective quotient of $(b^*_{-})^{-1}(1)$) is the space of time-like hyperplanes of Minkowski space.

- The boundary of an admissible convex set K in Minkowski space is the graph of a 1-Lipschitz convex function $u\colon \mathbb{R}^{n-1} \to \mathbb{R}$. From (3.5),

  $$\bar{h} = \sup_{(p,p_n)\in K} b_{n-1,1}\left(\begin{pmatrix} x \\ 1 \end{pmatrix}, \begin{pmatrix} p \\ p_n \end{pmatrix}\right)$$

  i.e.

  $$\bar{h} = \sup_{(p,p_n)\in K} \{b_{n-1,0}(x,p) - p_n\}$$

  i.e.

  $$\bar{h} = \sup_{p\in\mathbb{R}^d} \{b_{n-1,0}(x,p) - u(p)\},$$

  i.e. u is the dual of $\bar{h}$ for the Legendre–Fenchel duality, see e.g. [14, 2.3] and references therein for more details.

- By Fact 3.25, a negative convex function on the open unit ball B is the support function $\bar{h}$ of an admissible convex set K in Minkowski space. Let us suppose that the lower semi-continuous extension g of $\bar{h}$ to the boundary of B is continuous (this is not always the case, for example g could attain no maximum, see p. 870 of [33]). Let h_0 be the convex function on B such that the lower boundary of the convex hull of the graph of g is the graph of h_0. Then $h_0 \geq h$. The function h_0 is the support function of a convex subset of Minkowski space which is the *Cauchy development* of the convex set whose support function is h. The function g encodes the light-like support planes of K (and $\partial B \times \mathbb{R}$, the absolute of co-Minkowski space in a suitable affine model, is sometimes called the *Penrose boundary*). Note that a light-like support plane may not touch the boundary of K, as in the case of the upper branch of the hyperboloid.

Let h_0^+ be the concave function on B such that the upper boundary of the convex hull of the graph of g is the graph of h_0^+. It is possible to consider h_0^+ as the support function of a *past* convex set of Minkowski space. At the end of the day, at a projective level, it would be better to consider as "convex sets" of Minkowski space the data of a future convex set and a past convex set, with disjoint interior and same light-like support planes.

- The 3 dimensional spaces $\mathbb{H}^3$, $\mathbb{A}\mathrm{d}\mathbb{S}^3$ and $^*\mathbb{M}\mathrm{i}\mathbb{n}^3$ can be defined in a unified way as spaces of matrices with coefficient in $\mathbb{R} + \kappa\mathbb{R}$, with $\kappa \notin \mathbb{R}$, $\kappa^2 \in \mathbb{R}$, see [30].
- The reference [30] also contains a Gauss–Bonnet formula for the $^*\mathbb{M}\mathrm{i}\mathbb{n}^2$ geometry.

4 Geometric transition

In this section, we will study the so-called *geometric transition* of model spaces as subsets of projective space. Recall that a model space is an open subset $\mathbb{M}$ of $\mathbb{R}\mathrm{P}^n$, endowed with a closed subgroup $\mathrm{Isom}(\mathbb{M})$ of $\mathrm{PGL}(n+1, \mathbb{R})$ which preserves the geometric structure of $\mathbb{M}$.

Moreover, recall that given a model space $\mathbb{M}$, by applying a projective transformation $g \in \mathrm{PGL}(n+1, \mathbb{R})$ one obtains another model space which is equivalent to $\mathbb{M}$. Indeed, the group of isometries of $g\mathbb{M}$ is precisely $g\,\mathrm{Isom}(\mathbb{M})g^{-1}$.

We say that a model space $(\mathbb{N}, \mathrm{Isom}(\mathbb{N}))$ is a *conjugacy limit* or *rescaled limit* of another model space $(\mathbb{M}, \mathrm{Isom}(\mathbb{M}))$ if there exists a sequence of projective transformations $g_k \in \mathrm{PGL}(n+1, \mathbb{R})$ such that

i) the sequence $g_k\mathbb{M}$ converges to $\mathbb{N}$ as $k \to \infty$;

ii) the sequence of closed subgroups $g_k\,\mathrm{Isom}(\mathbb{M})g_k^{-1}$ converges to $\mathrm{Isom}(\mathbb{N})$.

The convergence here should be meant as the Hausdorff convergence, for instance. Thus the conditions i) and ii) essentially mean that

i′) every $x \in \mathbb{N}$ is the limit of a sequence $\{g_k x_k\}$, for some $x_k \in \mathbb{M}$, as $k \to \infty$;

ii′) every $h \in \mathrm{Isom}(\mathbb{N})$ is the limit of a sequence $\{g_k h_k g_k^{-1}\}$, for some $h_k \in \mathrm{Isom}(\mathbb{M})$, as $k \to \infty$.

In general, the sequences g_k and h_k are not compact sequences in $\mathrm{PGL}(n+1,\mathbb{R})$.

A toy model of a geometric transition is the 1-dimensional case, namely, the degeneration of projective lines. We already know that a line c in a model space is a subset of a copy of $\mathbb{R}\mathrm{P}^1$, with 0, 1 or 2 points in the absolute if c is elliptic, parabolic or hyperbolic respectively. More precisely, composing with a projective transformation, we can assume that c is $\mathbb{R}\mathrm{P}^1$ (if elliptic), $\mathbb{R}\mathrm{P}^1 \setminus \{[1:0]\}$ (parabolic) or $\mathbb{R}\mathrm{P}^1 \setminus \{[1:1],[1:-1]\}$ (hyperbolic). The stabilizer of c in the isometry group of the model space is identified with a subgroup of $\mathrm{PGL}(2,\mathbb{R})$ of the form

$$\left\{R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\right\}, \quad \left\{T_a = \begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix}\right\}, \quad \left\{S_\varphi = \begin{bmatrix} \cosh\varphi & -\sinh\varphi \\ \sinh\varphi & \cosh\varphi \end{bmatrix}\right\},$$

for c elliptic, parabolic of hyperbolic respectively. By applying the projective transformations

$$g_k = \begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix},$$

one sees that given a sequence θ_k, the sequence of conjugates

$$g_k R_{\theta_k} g_k^{-1} = \begin{bmatrix} \cos\theta_k & -k\sin\theta_k \\ (1/k)\sin\theta_k & \cos\theta_k \end{bmatrix}$$

converges to a projective transformation of the form T_a provided $\theta_k \sim a/k$. In an analogous way, $g_k S_{\varphi_k} g_k^{-1}$ converges to T_a as $k \to \infty$ if $\varphi_k \sim \frac{a}{k}$, and g_t maps the two points at infinity $[1:1]$ and $[1:-1]$ of the hyperbolic line to $\left[1:\frac{1}{k}\right]$ and $\left[1:-\frac{1}{k}\right]$, which converge to $[1:0]$ as $k \to \infty$. Hence:

Fact 4.1. *Elliptic and hyperbolic lines converge* (*in the sense of geometric transition*) *to a parabolic line.*

See also Figure 20. We will meet this phenomenon again in higher dimensions. We start by discussing some examples of geometric transition of 2-dimensional model spaces in $\mathbb{R}\mathrm{P}^3$.

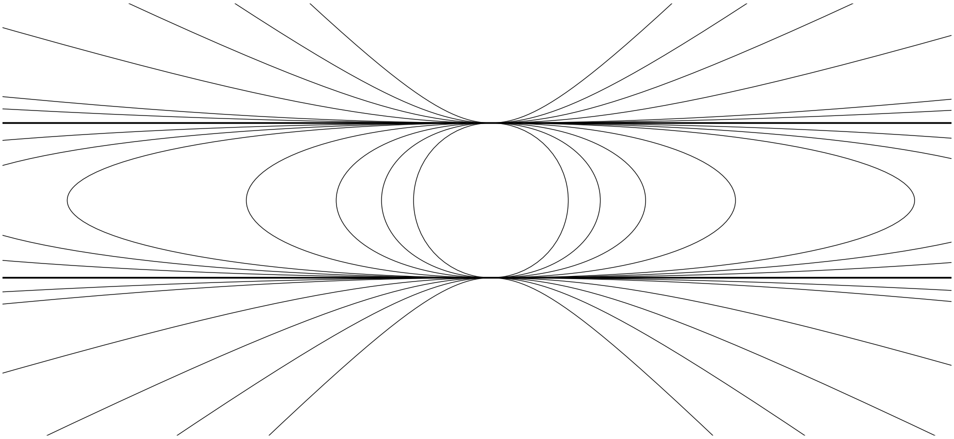


Figure 20. Geometric transition of elliptic line and hyperbolic line to Euclidean line, represented in the double cover

4.1 Limits of 2-dimensional model spaces

Recall that, in dimension 2, we have already introduced the model spaces corresponding to the elliptic plane $\mathbb{E}\mathbb{l}\mathbb{l}^2$, the hyperbolic plane $\mathbb{H}^2$, the de Sitter plane $\mathrm{d}\mathbb{S}^2$ (which is anti-isometric to $\mathbb{A}\mathrm{d}\mathbb{S}^2$) and the degenerate model spaces of the Euclidean and Minkowski planes $\mathbb{E}^2$ and $\mathbb{M}\mathbb{i}\mathbb{n}^2$, and the co-Euclidean and co-Minkowski planes $^*\mathbb{E}^2$ and $^*\mathbb{M}\mathbb{i}\mathbb{n}^2$.

Elliptic and hyperbolic planes limit to Euclidean plane. Let us consider the unit sphere $\mathcal{S}^2$ in the Euclidean space $\mathbb{E}^3$. We shall construct a *geometric transition* which has the Euclidean plane as a conjugacy limit. For this purpose, consider the following projective transformations $g_t \in \mathrm{GL}(3, \mathbb{R})$, for $t \in (0, 1]$:

$$g_t = \begin{pmatrix} \frac{1}{t} & 0 & 0 \\ 0 & \frac{1}{t} & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The transformations g_t map the unit sphere $\mathcal{S}^2$ to an ellipsoid.

Observe that g_t fixes the point $(0,0,1) \in \mathbb{E}^3$ and "stretches" the directions x_1, x_2. More precisely, given a differentiable path $x(t) \in \mathbb{S}^2$, for $t \in [0,1]$, such that $x(0) = (0,0,1)$, one has

$$\lim_{t \to 0} g_t x_t = \lim_{t \to 0} \begin{pmatrix} \dfrac{x_1(t)}{t} \\ \dfrac{x_2(t)}{t} \\ x_3(t) \end{pmatrix} = \begin{pmatrix} \dot{x}_1(0) \\ \dot{x}_2(0) \\ 1 \end{pmatrix}.$$

Roughly speaking, the rescaled limit is a point of the affine chart $\{x_3 = 1\}$, which encodes the first order derivative of the x_1, x_2 coordinates. Heuristically, this rescaling procedure is a *blow-up of a point*, in the sense that the point $(0,0,1)$ is preserved, and the transverse directions are "blown-up."

This procedure is actually well defined in projective space. Indeed the transformations g_t descend to projective transformations of $\mathbb{R}\mathrm{P}^2$, which we will still denote by g_t. The points of the Euclidean plane $\mathbb{E}^2$, which is defined as

$$\mathbb{E}^2 = \{x \in \mathbb{E}^3 \mid x_3^2 = 1\}/\{\pm \mathrm{Id}\}$$

are conjugacy limits as $t \to 0$ of sequences of points of the elliptic plane

$$\mathbb{E}\mathbb{l}\mathbb{l}^2 = \{x \in \mathbb{E}^3 \mid b_{3,0}(x,x) = 1\}/\{\pm \mathrm{Id}\},$$

thus satisfying condition i) in the definition of conjugacy limit (see p. 359), for instance with $t = 1/k$, see Figure 21. To check that the Euclidean plane is a conjugacy limit of the elliptic space, one has to check also condition ii) on the isometry groups. That is, given an isometry of $\mathbb{E}\mathbb{l}\mathbb{l}^2$, namely an element in PO(3), let us choose a representative with determinant ± 1 of the form

$$h = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{bmatrix}.$$

By a direct computation, one checks that

$$g_t h g_t^{-1} = \begin{bmatrix} a_{11} & a_{12} & \dfrac{a_{13}}{t} \\ a_{12} & a_{22} & \dfrac{a_{23}}{t} \\ t a_{13} & t a_{23} & a_{33} \end{bmatrix}.$$

In fact, $g_t h g_t^{-1}$ is a transformation which preserves the quadratic form $\left(\frac{1}{t}\right)^2 x_1^2 + \left(\frac{1}{t}\right)^2 x_2^2 + x_3^2$. Thus every accumulation point of sequences of the form $g_t h(t) g_t^{-1}$ is necessarily of the form

$$h_\infty = \begin{bmatrix} a_{11}(0) & a_{12}(0) & \dot{a}_{13}(0) \\ a_{12}(0) & a_{22}(0) & \dot{a}_{23}(0) \\ 0 & 0 & \pm 1 \end{bmatrix} \in \mathrm{Isom}(\mathbb{E}^2),$$

where

$$\begin{pmatrix} a_{11}(0) & a_{12}(0) \\ a_{12}(0) & a_{22}(0) \end{pmatrix} \in \mathrm{O}(2).$$

and

$$h(0) = \begin{bmatrix} a_{11}(0) & a_{12}(0) & 0 \\ a_{12}(0) & a_{22}(0) & 0 \\ 0 & 0 & \pm 1 \end{bmatrix}.$$

This shows that the conjugacy limit of the group of isometries of $\mathbb{E}\mathbb{l}\mathbb{l}^2$ is precisely the group of isometries of $\mathbb{E}^2$, embedded in PGL$(3, \mathbb{R})$ in the usual way (see (8) and (9)). By a completely analogous proof, using the same transformations g_t, one sees that $\mathbb{E}^2$ is a conjugacy limit of the hyperbolic plane $\mathbb{H}^2$, see Figure 21. Thus one can imagine that the Euclidean plane (at $t = 0$) is an interpolation of the elliptic plane (for $t > 0$) and the hyperbolic plane (for $t < 0$).

Elliptic and de Sitter plane limit to co-Euclidean plane. We now describe a different procedure which permits to obtain a different limit from the elliptic plane, namely, we will obtain the co-Euclidean plane as a conjugacy limit of the elliptic plane. Thus, consider

$$g_t^* = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{t} \end{bmatrix} \in \mathrm{PGL}(3, \mathbb{R})$$

for $t \in (0, 1]$. As a remarkable difference with the previous case, the projective transformation g_t^* leaves invariant a geodesic line of $\mathbb{E}\mathbb{l}\mathbb{l}^2$, namely the line which is defined by the plane $\{x_3 = 0\}$ of $\mathbb{E}^3$. So the "stretching" occurs only in the transverse directions to $\{x_3 = 0\}$, and indeed the rescaled limit of a differentiable path of points of the form $x(t) = [x_1(t) : x_2(t) : x_3(t)]$ such that $x_3(0) = 0$ is

$$\lim_{t \to 0} g_t^* x_t = \lim_{t \to 0} \begin{bmatrix} x_1(t) \\ x_2(t) \\ \frac{x_3(t)}{t} \end{bmatrix} = \begin{bmatrix} x_1(0) \\ x_2(0) \\ \dot{x}_3(0) \end{bmatrix}.$$

We will indeed call this transition the *blow-up of a line*. This shows that points of the co-Euclidean plane, defined by

$$^*\mathbb{E}^2 = \{x \in \mathbb{E}^3 \mid b^*(x,x) = 1\}/\{\pm \mathrm{Id}\}$$

(where $b^*(x,x) = x_1^2 + x_2^2$) are rescaled limits of sequences in

$$\mathbb{Ell}^2 = \{x \in \mathbb{E}^3 \mid b_{3,0}(x,x) = 1\}/\{\pm \mathrm{Id}\}.$$

For what concerns the isometry groups, we will give again a computation which shows that $\mathrm{Isom}(^*\mathbb{E}^2)$ is the limit of $g_t^* \,\mathrm{Isom}(\mathbb{Ell}^2)(g_t^*)^{-1}$. As before, choose a representative of determinant ±1, say

$$h = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{bmatrix} \in \mathrm{PGL}(3,\mathbb{R}).$$

By a direct computation,

$$g_t^* h (g_t^*)^{-1} = \begin{bmatrix} a_{11} & a_{12} & t a_{13} \\ a_{12} & a_{22} & t a_{23} \\ \dfrac{a_{13}}{t} & \dfrac{a_{23}}{t} & a_{33} \end{bmatrix} \longrightarrow \begin{bmatrix} a_{11}(0) & a_{12}(0) & 0 \\ a_{12}(0) & a_{22}(0) & 0 \\ \dot{a}_{13}(0) & \dot{a}_{23}(0) & \pm 1 \end{bmatrix},$$

provided $a_{13}(0) = a_{23}(0) = 0$ and

$$\begin{pmatrix} a_{11}(0) & a_{12}(0) \\ a_{12}(0) & a_{22}(0) \end{pmatrix} \in \mathrm{O}(2).$$

In conclusion the limit of $g_t^* h (g_t^*)^{-1}$ is an element of the group $\mathrm{Isom}(^*\mathbb{E}^2)$ in Definition 3.10.

Recall that we have introduced a duality of $\mathbb{Ell}^2$ to itself, which maps lines of $\mathbb{Ell}^2$ to points of $\mathbb{Ell}^2$ and vice versa. In principle, the dual space is a model space in the dual projective space. The dual space of $\mathbb{Ell}^2$ is thus identified to $\mathbb{Ell}^2$ itself if one chooses the scalar product of $\mathbb{E}^3$ to identify $\mathbb{R}\mathrm{P}^2$ to its dual projective space. Now, observe that the transformations g_t and g_t^*, which have been used to rescale $\mathbb{Ell}^2$ to get $\mathbb{E}^2$ and $^*\mathbb{E}^2$ respectively, are well-behaved with respect to this duality. Namely, $b_{3,0}(x,y) = 0$ if and only if $b_{3,0}(g_t x, g_t^* y) = 0$. In other words, if $*$ denotes the duality point-line induced by the ambient scalar product, the following diagram is commutative:

$$\begin{tikzcd}
(\mathbb{Ell}^2, \mathrm{Isom}(\mathbb{Ell}^2)) \arrow[r, "g_t"] \arrow[d, leftrightarrow, "*"] & (g_t \mathbb{Ell}^2, g_t \,\mathrm{Isom}(\mathbb{Ell}^2) g_t^{-1}) \arrow[d, leftrightarrow, "*"] \\
(\mathbb{Ell}^2, \mathrm{Isom}(\mathbb{Ell}^2)) \arrow[r, "g_t^*"] & (g_t^* \mathbb{Ell}^2, g_t^* \,\mathrm{Isom}(\mathbb{Ell}^2)(g_t^*)^{-1})
\end{tikzcd}$$

When $t \to 0$, we have already observed that $(g_t \mathbb{E}\mathrm{ll}^2, g_t \operatorname{Isom}(\mathbb{E}\mathrm{ll}^2) g_t^{-1})$ converges to $(\mathbb{E}^2, \operatorname{Isom}(\mathbb{E}^2))$, while on the other hand $(g_t^* \mathbb{E}\mathrm{ll}^2, g_t^* \operatorname{Isom}(\mathbb{E}\mathrm{ll}^2)(g_t^*)^{-1})$ converges to $({}^*\mathbb{E}^2, {}^* \operatorname{Isom}({}^*\mathbb{E}^2))$. The natural duality of $\mathbb{E}^2$ and ${}^*\mathbb{E}^2$ is again induced by the scalar product of $\mathbb{E}^3$ and the commutativity of the diagram passes to the limit. Thus, we have shown:

Fact 4.2. *The dual in* ${}^*\mathbb{E}^2$ *of a rescaled limit* $x_\infty \in \mathbb{E}^2$ *of points* $x(t) \in \mathbb{E}\mathrm{ll}^2$ *is the rescaled limit of the dual lines* $x(t)^*$ *in* $\mathbb{E}\mathrm{ll}^2$ *and vice versa (by exchanging the roles of points and lines).*

In a completely analogous way, one can define a geometric transition which permits to obtain the co-Euclidean plane ${}^*\mathbb{E}^2$ as a conjugacy limit of the de Sitter plane $\mathrm{d}\mathbb{S}^2$. Indeed, observe that $\mathbb{E}\mathrm{ll}^2$ and $\mathrm{d}\mathbb{S}^2$ are the constant curvature non-degenerate model spaces in dimension 2 which contain an elliptic line $\mathbb{E}\mathrm{ll}^1$, whereas ${}^*\mathbb{E}^2$ is the degenerate space having an embedded copy of $(\mathbb{E}\mathrm{ll}^1, \operatorname{Isom}(\mathbb{E}\mathrm{ll}^1))$. Therefore the following fact holds:

Fact 4.3. *The dual in* ${}^*\mathbb{E}^2$ *of a rescaled limit* $x_\infty \in \mathbb{E}^2$ *of points* $x(t) \in \mathbb{H}^2$ *is the rescaled limit of the dual lines* $x(t)^*$ *in* $\mathrm{d}\mathbb{S}^2$*. The dual in* $\mathbb{E}^2$ *of a rescaled limit* $x_\infty \in {}^*\mathbb{E}^2$ *of points* $x(t) \in \mathrm{d}\mathbb{S}^2$ *is the rescaled limit of the dual lines* $x(t)^*$ *in* $\mathbb{H}^2$*. The converse is also true, by exchanging the role of points and lines.*

We thus have the following diagram, which encodes the possible transitions and dualities involving $\mathbb{E}^2$ and ${}^*\mathbb{E}^2$, see Figures 21 and 22:

$$\begin{tikzcd}
(\mathbb{E}\mathrm{ll}^2, \operatorname{Isom}(\mathbb{E}\mathrm{ll}^2)) \arrow[r, "\text{blow-up point}"] \arrow[d, leftrightarrow, "*"] & (\mathbb{E}^2, \operatorname{Isom}(\mathbb{E}^2)) \arrow[d, leftrightarrow, "*"] & (\mathbb{H}^2, \operatorname{Isom}(\mathbb{H}^2)) \arrow[l, "\text{blow-up point}"'] \arrow[d, leftrightarrow, "*"] \\
(\mathbb{E}\mathrm{ll}^2, \operatorname{Isom}(\mathbb{E}\mathrm{ll}^2)) \arrow[r, "\text{blow-up line}"] & ({}^*\mathbb{E}^2, \operatorname{Isom}({}^*\mathbb{E}^2)) & (\mathrm{d}\mathbb{S}^2, \operatorname{Isom}(\mathrm{d}\mathbb{S}^2)) \arrow[l, "\text{blow-up line}"']
\end{tikzcd} \tag{22}$$

Geometric transitions with limits Minkowski and co-Minkowski planes. Given the above constructions, it is immediate to see that one can mimic the blow-up of a point for the de Sitter plane

$$\mathrm{d}\mathbb{S}^2 = \{x \in \mathbb{M}\mathrm{in}^3 \mid b_{2,1}(x, x) = 1\}/\{\pm \mathrm{Id}\}.$$

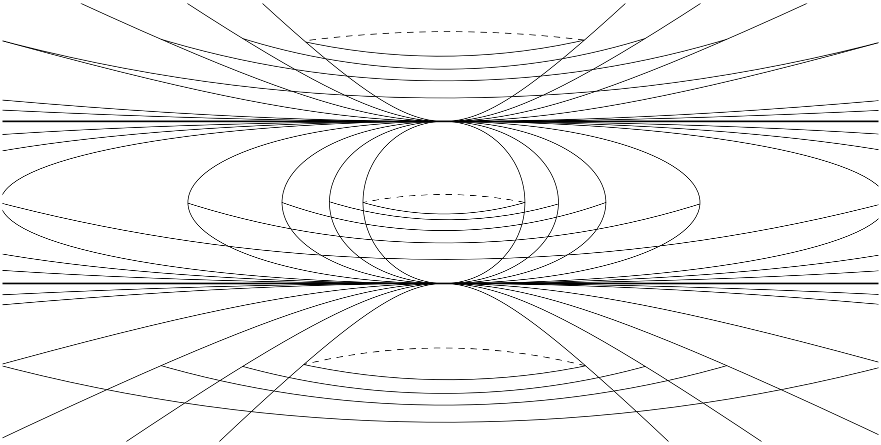


Figure 21. Degenerations of the top line in Scheme (22), inside the ambient space, in the double cover

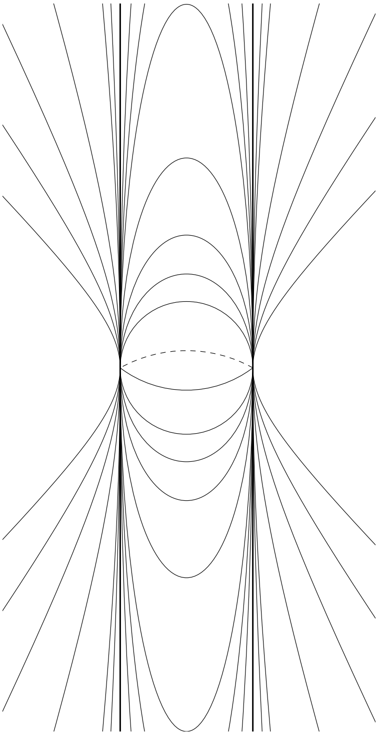


Figure 22. Degenerations of the bottom line in Scheme (22), inside the ambient space, in the double cover

Using again the transformations g_t, the limit will be again represented by an affine chart defined by $\{x_3^2 = 1\}/\{\pm \mathrm{Id}\}$, and the conjugacy limit of $\mathrm{Isom}(\mathrm{d}\mathbb{S}^2) \cong \mathrm{PO}(2,1)$ will be

$$\mathrm{Isom}(\mathbb{Mim}^2) = \left\{ \begin{bmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ 0 & 0 & \pm 1 \end{bmatrix} : \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix} \in \mathrm{O}(1,1) \right\}.$$

Hence we can say that the blow-up of a point permits to obtain $(\mathbb{Mim}^2, \mathrm{Isom}(\mathbb{Mim}^2))$ as a limit of $(\mathrm{d}\mathbb{S}^2, \mathrm{Isom}(\mathrm{d}\mathbb{S}^2))$.

Also in this case we shall obtain the dual transition. As in the case of the Euclidean/co-Euclidean plane, there are two possible transitions having limit the co-Minkowski plane. Indeed, both the hyperbolic plane and the de Sitter plane contain hyperbolic lines (space-like in $\mathbb{H}^2$, time-like in $\mathrm{d}\mathbb{S}^2$). For the hyperbolic plane, one checks directly that

$$\begin{aligned} & g_t^* \colon \mathrm{Isom}(\mathbb{H}^2)(g_t^*)^{-1} \\ & \quad \longrightarrow \mathrm{Isom}({}^*\mathbb{Mim}^2) = \left\{ \begin{bmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ v_1 & v_2 & \pm 1 \end{bmatrix} : \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix} \in \mathrm{O}(1,1) \right\}. \end{aligned}$$

Observe that, in terms of geometric transition, if $(\mathbb{N}, \mathrm{Isom}(\mathbb{N}))$ is a conjugacy limit of $(\mathbb{M}, \mathrm{Isom}(\mathbb{M}))$, then also the anti-isometric space $(\overline{\mathbb{N}}, \mathrm{Isom}(\overline{\mathbb{N}}))$ is a limit of $(\mathbb{M}, \mathrm{Isom}(\mathbb{M}))$ (and also of $(\overline{\mathbb{M}}, \mathrm{Isom}(\overline{\mathbb{M}}))$, of course), by conjugating for the same projective transformations. For instance, both $\mathrm{d}\mathbb{S}^2$ and $\overline{\mathrm{d}\mathbb{S}^2}$ limit to ${}^*\mathbb{Mim}^2$. However, we prefer to say that $(\overline{\mathrm{d}\mathbb{S}^2}, \mathrm{Isom}(\overline{\mathrm{d}\mathbb{S}^2}))$ limits to $({}^*\mathbb{Mim}^2, \mathrm{Isom}({}^*\mathbb{Mim}^2))$ by blowing-up a space-like line, since with this choice, space-like lines (of hyperbolic type) converge to space-like lines of ${}^*\mathbb{Mim}^2$.

For the same reason, blowing-up a point one gets that $\overline{\mathrm{d}\mathbb{S}^2}$ limits to $\overline{\mathbb{Mim}^2}$. However, in this special case, the space $\overline{\mathbb{Mim}^2}$, anti-isometric to $\mathbb{Mim}^2$, is also isometric to $\mathbb{Mim}^2$, thus we have a geometric transition

$$(\overline{\mathrm{d}\mathbb{S}^2}, \mathrm{Isom}(\overline{\mathrm{d}\mathbb{S}^2})) \longrightarrow (\mathbb{Mim}^2, \mathrm{Isom}(\mathbb{Mim}^2)).$$

This is formally not the same as the transition of $\mathrm{d}\mathbb{S}^2$ to $\mathbb{Mim}^2$ (in $\mathrm{d}\mathbb{S}^2$ we have space-like elliptic lines which converge to space-like lines of $\mathbb{Mim}^2$, while in $\overline{\mathrm{d}\mathbb{S}^2}$ the space-like lines are hyperbolic and converge to space-like lines), although it is obtained for instance by applying the usual transformations g_t. Recall also that the dual of $\overline{\mathrm{d}\mathbb{S}^2}$, considered as the space of space-like lines, is $\overline{\mathrm{d}\mathbb{S}^2}$ itself (see Figure 7).

By the same argument as in the previous paragraph (using the ambient metric of $\mathbb{M}\mathrm{in}^3$ on the left, and its anti-isometric $\overline{\mathbb{M}\mathrm{in}^3}$ on the right), one obtains the following diagram which shows the behavior of transitions and dualities, see Figure 23:

$$\begin{tikzcd}
(\mathrm{d}\mathbb{S}^2, \mathrm{Isom}(\mathrm{d}\mathbb{S}^2)) \arrow[r, "\text{blow-up point}"] \arrow[d, leftrightarrow, "*"] & (\mathbb{M}\mathrm{in}^2, \mathrm{Isom}(\mathbb{M}\mathrm{in}^2)) \arrow[d, leftrightarrow, "*"] & (\overline{\mathrm{d}\mathbb{S}^2}, \mathrm{Isom}(\overline{\mathrm{d}\mathbb{S}^2})) \arrow[l, "\text{blow-up point}"'] \arrow[d, leftrightarrow, "*"] \\
(\mathbb{H}^2, \mathrm{Isom}(\mathbb{H}^2)) \arrow[r, "\text{blow-up line}"] & ({}^*\mathbb{M}\mathrm{in}^2, \mathrm{Isom}({}^*\mathbb{M}\mathrm{in}^2)) & (\overline{\mathrm{d}\mathbb{S}^2}, \mathrm{Isom}(\overline{\mathrm{d}\mathbb{S}^2})) \arrow[l, "\text{blow-up line}"']
\end{tikzcd} \tag{23}$$

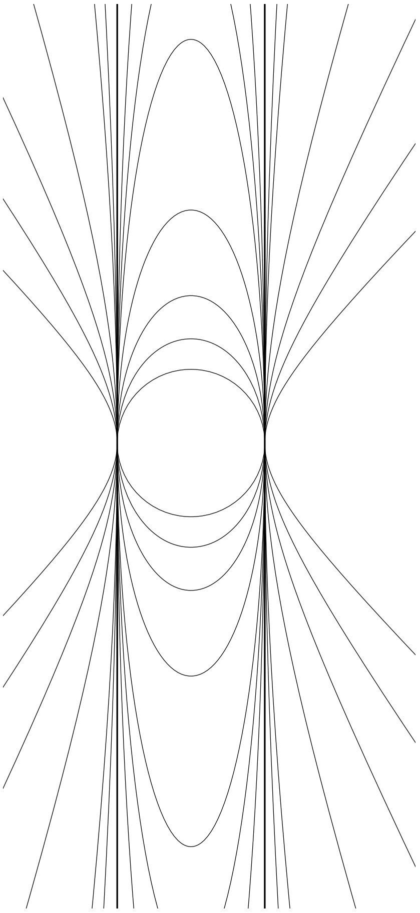


Figure 23. Degenerations of the bottom line in Scheme (23), in an affine chart (recall that $\overline{\mathrm{d}\mathbb{S}^2}$ is the same as $\mathbb{A}\mathrm{d}\mathbb{S}^2$)

In words:

Fact 4.4. *The dual in* $^*\mathbb{Min}^2$ *of a rescaled limit* $x_\infty \in \mathbb{Min}^2$ *of points* $x(t) \in \mathrm{d}\mathbb{S}^2$ *is the rescaled limit of the dual lines* $x(t)^*$ *in* $\mathbb{H}^2$. *The dual in* $\mathbb{Min}^2$ *of a rescaled limit* $x_\infty \in {}^*\mathbb{Min}^2$ *of points* $x(t) \in \mathbb{H}^2$ *is the rescaled limit of the dual lines* $x(t)^*$ *in* $\mathrm{d}\mathbb{S}^2$. *The converse is also true, by exchanging the role of points and lines.*

Fact 4.5. *The dual in* $^*\mathbb{Min}^2$ *of a rescaled limit* $x_\infty \in \mathbb{Min}^2$ *of points* $x(t) \in \overline{\mathrm{d}\mathbb{S}^2}$ *is the rescaled limit of the dual lines* $x(t)^*$ *in* $\overline{\mathrm{d}\mathbb{S}^2}$ *and vice versa* (*by exchanging the roles of points and lines*).

4.2 Limits of 3-dimensional model spaces

At this stage, the reader will not be surprised to find that the transition procedures described in the previous paragraph extend also to the three-dimensional case (and to higher dimensions, although this will not be considered in this survey). For instance, the following diagram summarizes the transitions which have limits in the Euclidean space or the co-Euclidean space:

$$\begin{tikzcd}
(\mathbb{Ell}^3, \mathrm{Isom}(\mathbb{Ell}^3)) \arrow[r, "\text{blow-up point}"] \arrow[d, leftrightarrow, "*"] & (\mathbb{E}^3, \mathrm{Isom}(\mathbb{E}^3)) \arrow[d, leftrightarrow, "*"] & (\mathbb{H}^3, \mathrm{Isom}(\mathbb{H}^3)) \arrow[l, "\text{blow-up point}"'] \arrow[d, leftrightarrow, "*"] \\
(\mathbb{Ell}^3, \mathrm{Isom}(\mathbb{Ell}^3)) \arrow[r, "\text{blow-up plane}"] & ({}^*\mathbb{E}^3, \mathrm{Isom}({}^*\mathbb{E}^3)) & (\mathrm{d}\mathbb{S}^3, \mathrm{Isom}(\mathrm{d}\mathbb{S}^3)) \arrow[l, "\text{blow-up plane}"']
\end{tikzcd} \tag{24}$$

Indeed, the way to rescale elliptic space or de Sitter space to get a limit in co-Euclidean space is by *blowing-up a plane*. Space-like planes in $\mathbb{Ell}^3$ and $\mathrm{d}\mathbb{S}^3$ are indeed copies of $\mathbb{Ell}^2$, and $^*\mathbb{E}^3$ is the degenerate geometry having an embedded $\mathbb{Ell}^2$ plane. It is thus an exercise to rewrite the statements of Facts 4.2 and 4.3 in the three-dimensional setting, by making use of the duality points/planes.

The other diagram we considered in dimension 2 also has a generalization here. In fact, it will now become clear that the right model spaces which have a limit in Minkowski space (by blowing-up a point) are de Sitter and anti-de Sitter space; whereas their duals (hyperbolic space and anti-de Sitter space itself) have a limit in co-Minkowski space, see Figure 24:

$$\begin{tikzcd}
(\mathrm{d}\mathbb{S}^3, \mathrm{Isom}(\mathrm{d}\mathbb{S}^3)) \arrow[r, "\text{blow-up point}"] \arrow[d, leftrightarrow, "*"] & (\mathbb{Min}^3, \mathrm{Isom}(\mathbb{Min}^3)) \arrow[d, leftrightarrow, "*"] & (\mathbb{A}\mathrm{d}\mathbb{S}^3, \mathrm{Isom}(\mathbb{A}\mathrm{d}\mathbb{S}^3)) \arrow[l, "\text{blow-up point}"'] \arrow[d, leftrightarrow, "*"] \\
(\mathbb{H}^3, \mathrm{Isom}(\mathbb{H}^3)) \arrow[r, "\text{blow-up plane}"] & ({}^*\mathbb{Min}^3, \mathrm{Isom}({}^*\mathbb{Min}^3)) & (\mathbb{A}\mathrm{d}\mathbb{S}^3, \mathrm{Isom}(\mathbb{A}\mathrm{d}\mathbb{S}^3)) \arrow[l, "\text{blow-up plane}"']
\end{tikzcd} \tag{25}$$

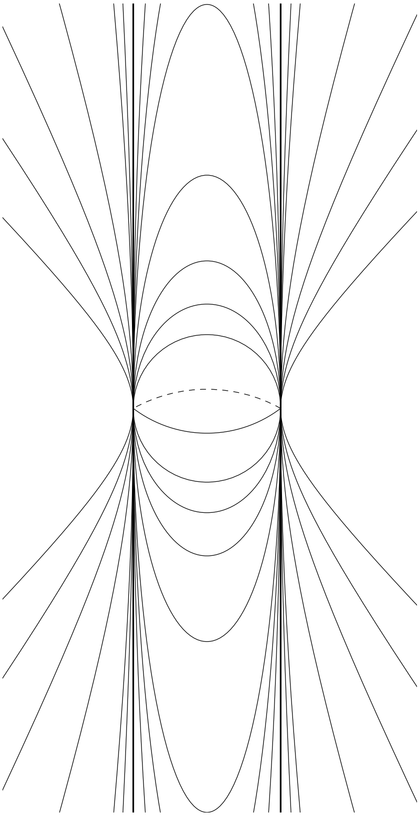


Figure 24. Degenerations of the bottom line in Scheme (25), in an affine chart

In fact, in dimension 2 we did not define an anti-de Sitter plane, as it would be anti-isometric to the de Sitter plane. Thus we write here the statement which encodes the relationship between transitions and dualities for anti-de Sitter space, leaving the analogous statements to the reader:

Fact 4.6. *The dual in* $^*\mathbb{M}\mathrm{i}\mathbb{n}^3$ *of a rescaled limit* $x_\infty \in \mathbb{M}\mathrm{i}\mathbb{n}^3$ *of points* $x(t) \in \mathbb{A}\mathrm{d}\mathbb{S}^3$ *is the rescaled limit of the dual planes* $x(t)^*$ *in* $\mathbb{A}\mathrm{d}\mathbb{S}^3$ *and vice versa (by exchanging the roles of points and planes).*

Blowing-up time-like planes. However, we observe that in dimension three there are more complicated pheonomena which might occur. In particular, one might be interested in the blow-up of a time-like plane in a Lorentzian model-space. Of course this is the same of blowing-up a space-like plane in the anti-isometric model space. Intrinsically, the time-like plane can be a copy of $\mathrm{d}\mathbb{S}^2$ or $\overline{\mathrm{d}\mathbb{S}^2}$. Let us analyze the following diagram:

$$\begin{tikzcd}
(\overline{\mathrm{d}\mathbb{S}^3}, \mathrm{Isom}(\overline{\mathrm{d}\mathbb{S}^3})) \arrow[r, "\text{blow-up point}"] \arrow[d, leftrightarrow, "*"] & (\overline{\mathbb{M}\mathrm{in}^3}, \mathrm{Isom}(\overline{\mathbb{M}\mathrm{in}^3})) \arrow[d, leftrightarrow, "*"] & (\overline{\mathbb{A}\mathrm{d}\mathbb{S}^3}, \mathrm{Isom}(\overline{\mathbb{A}\mathrm{d}\mathbb{S}^3})) \arrow[l, "\text{blow-up point}"'] \arrow[d, leftrightarrow, "*"] \\
(\mathrm{d}\mathbb{S}^3, \mathrm{Isom}(\mathrm{d}\mathbb{S}^3)) \arrow[r, "\text{blow-up plane}"] & (^*\overline{\mathbb{M}\mathrm{in}^3}, \mathrm{Isom}(^*\overline{\mathbb{M}\mathrm{in}^3})) & (\overline{\mathbb{A}\mathrm{d}\mathbb{S}^3}, \mathrm{Isom}(\overline{\mathbb{A}\mathrm{d}\mathbb{S}^3})) \arrow[l, "\text{blow-up plane}"']
\end{tikzcd}$$

On the upper line, there is nothing surprising. We have just re-written the usual blow-up of a point, but by considering the anti-isometric copies of $\mathrm{d}\mathbb{S}^3$, $\mathbb{M}\mathrm{in}^3$ and $\mathbb{A}\mathrm{d}\mathbb{S}^3$. In fact, the definition of geometric transition does not distinguish between a model space and its anti-isometric copy. Anyway, we decided to stick to the convention to choose the sign of the metric in such a way that space-like lines converge to space-like lines, and so on.

On the lower line, we have already encountered the duality which appears on the left. In fact, in $\mathbb{M}\mathrm{in}^4$ the linear hyperplanes which define time-like planes in $\mathrm{d}\mathbb{S}^3$ also define planes in $\mathbb{H}^3$ (see Remark 2.18). Hence the space of time-like planes in $\mathrm{d}\mathbb{S}^3$, or equivalently the space of space-like planes in $\overline{\mathrm{d}\mathbb{S}^3}$, is naturally $\mathrm{d}\mathbb{S}^3$ itself.

In the right-hand side, recall that $\overline{\mathbb{A}\mathrm{d}\mathbb{S}^3}$ can be defined as

$$\overline{\mathbb{A}\mathrm{d}\mathbb{S}^3} = \{x \in \mathbb{R}^{2,2} \mid b_{2,2}(x) = 1\}/\{\pm \mathrm{Id}\}.$$

Thus a space-like plane in $\overline{\mathbb{A}\mathrm{d}\mathbb{S}^3}$ is a copy of $\mathrm{d}\mathbb{S}^2$, and the dual space of $\overline{\mathbb{A}\mathrm{d}\mathbb{S}^3}$, considered as the space of space-like planes of $\overline{\mathbb{A}\mathrm{d}\mathbb{S}^3}$, is $\overline{\mathbb{A}\mathrm{d}\mathbb{S}^3}$ itself.

It remains to understand what is the limit space in the centre of the lower line. We denoted it by $^*\overline{\mathbb{M}\mathrm{in}^3}$ to indicate that it is the dual of $\overline{\mathbb{M}\mathrm{in}^3}$ (and not the space $\overline{^*\mathbb{M}\mathrm{in}^3}$ anti-isometric to $^*\mathbb{M}\mathrm{in}^3$!) We will omit the details of the definition and the proof of the commutativity of the last diagram presented. However, observe that in the lower line, both $\mathrm{d}\mathbb{S}^3$ and $\overline{\mathbb{A}\mathrm{d}\mathbb{S}^3}$ contain a totally geodesic copy of $\mathrm{d}\mathbb{S}^2$. Thus one can define a transition procedure which *blows-up a time-like plane*, stretching the transverse directions. Topologically the limit space is expected to be $\mathrm{d}\mathbb{S}^2 \times \mathbb{R}$. By a construction similar to that of $^*\mathbb{M}\mathrm{in}^3$, one can identify this space to the space of time-like planes in $\mathbb{M}\mathrm{in}^3$ (or of space-like planes in $\overline{\mathbb{M}\mathrm{in}^3}$). In an affine chart, this would be the exterior of the cylinder which represents $^*\mathbb{M}\mathrm{in}^3$.

Comments and references

- Let us denote by $\mathbb{G}\mathrm{al}^2$ the *Galilean plane*, i.e. the projective plane minus a parabolic line. See [69]. Below are shown the possible degenerations of the three Riemannian and the three Lorentzian plane geometries. This diagram is the one in Section 5.3 of [24], adapted to our terminology.

$$\begin{tikzcd}
\mathbb{Ell}^2 \arrow[d] \arrow[drr] & \mathbb{H}^2 \arrow[dl] \arrow[d] & \overline{\mathrm{d}\mathbb{S}^2} \arrow[dl] \arrow[d] \arrow[dr] & \\
\mathbb{E}^2 \arrow[dr] & {}^*\mathbb{Min}^2 \arrow[d] & {}^*\mathbb{E}^2 \arrow[dl] & \mathbb{Min}^2 \arrow[dll] \\
 & \mathbb{Gal}^2 & &
\end{tikzcd}$$

This essentially shows that the geometric transitions considered in this paper are all the possible transitions in $\mathbb{R}\mathrm{P}^3$, except the further space $\mathbb{Gal}^2$ which is, in some vague sense, doubly-degenerate. Essentially, the isometry group for $\mathbb{Gal}^2$ is the subset of triangular matrices which preserve an affine chart.

- In [24], the possible degenerations are classified in every dimension. Already in dimension 3, apart from those considered above, there are other *simple* degenerations, essentially obtained by blowing up a line. These spaces are described as model spaces $(X, \mathrm{Isom}(X))$ and in some cases they contain as subgeometries (i.e. as a geometry (X, G) where G is a subgroup of $\mathrm{Isom}(X)$) other 3-dimensional geometries in the sense of Thurston. For instance Sol geometry is a possible limit of hyperbolic structures [42]. One could draw the corresponding diagram in dimension 3, which would already be pretty complicated, and find several other spaces as double degenerations. For example, an affine space endowed with the action of a group of lower triangular matrices, with unitary elements on the diagonal, is a generalization of $\mathbb{Gal}^2$.

- For the effect of the two-dimensional transition hyperbolic-Euclidean-spherical, the effects on angles, area etc. of triangles are studied in [3].

5 Connection and volume form

In this section we discuss the definitions of the Levi-Civita connection and volume form for model spaces, starting by the general setting of Riemannian or pseudo-Riemannian manifolds and then specializing to the cases of constant curvature manifolds. We will then give a construction of a geometric connection and a volume form on the degenerate cases of co-Euclidean and co-Minkowski space.

5.1 Non-degenerate model spaces

Let (M, g) be a three-dimensional Riemannian or pseudo-Riemannian manifold. Many key examples have already been introduced: $\mathbb{E}^3$, $\mathbb{H}^3$ and $\mathbb{Ell}^3$ are Riemannian, $\mathbb{Min}^3$, $\mathrm{d}\mathbb{S}^3$ and $\mathbb{A}\mathrm{d}\mathbb{S}^3$ are pseudo-Riemannian. Recall that the non-degenerate metric g uniquely determines the *Levi-Civita connection*, which we will denote by ∇^g or ∇^M if there is no ambiguity about the metric tensor g. Indeed, the Levi-Civita connection is the unique connection which satisfies the following properties:

- ∇^g is *symmetric,* i.e. $\nabla^g_X Y - \nabla^g_Y X = [X, Y]$ for any pair of vector fields X, Y;
- ∇^g is *compatible with the metric,* i.e. $Z.g(X, Y) = g(\nabla^g_Z X, Y) + g(X, \nabla^g_Z Y)$ for any vector fields X, Y, Z.

The latter condition can also be expressed by saying that $\nabla^g g = 0$, namely the metric tensor is parallel with respect to ∇^g. Recall also that a *volume form* is defined on (M, g), provided M is oriented. Namely, the volume form is a 3-form ω such that $\omega(X, Y, Z) = 1$ if X, Y, Z is an oriented (with respect to the orientation of M) orthonormal frame for g. The volume form is also characterized by being the unique 3-form w such that

- ω is *parallel* with respect to ∇^g, i.e. $\nabla^g \omega = 0$;
- $\omega(v_1, v_2, v_3) = 1$, where v_1, v_2, v_3 is an oriented orthonormal triple of vectors at a fixed point x_0.

Indeed, as the parallel transport preserves an oriented orthonormal basis, the volume form is parallel for the connection, i.e. $\nabla^g \omega = 0$, i.e.

$$Z.\omega(X_1, \dots, X_n) = \omega(\nabla^g_Z X_1, \dots, X_n) + \cdots + \omega(X_1, \dots, \nabla^g_Z X_n). \qquad (26)$$

Explicit construction of the ambient Levi-Civita connection and volume form. Let us take a step back, and see how the Levi-Civita connection of the three-dimensional manifolds considered here can be defined. For instance, the Levi-Civita connection of Euclidean space is simply given by differentiation of the standard coordinates of a vector field, that is,

$$\nabla^{\mathbb{E}^3}_v w = Dw(v),$$

for any pair of smooth vector fields v, w. The connection of Minkowski space is defined analogously, and thus on the same affine space, the Levi-Civita connections of the Euclidean and the Minkowski metric coincide. Moreover, the standard volume form of $\mathbb{R}^3$ coincides with the volume forms (induced by the metric) of $\mathbb{E}^3$ and $\mathbb{M}\mathrm{in}^3$, i.e. in the standard coordinates (x, y, z):

$$\omega_{\mathbb{E}^3} = \omega_{\mathbb{M}\mathrm{in}^3} = \mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z = \omega_{\mathbb{R}^3}. \tag{27}$$

Of course, the same definitions can be given for any n-dimensional vector space endowed with a non-degenerate symmetric bilinear form b. Now we will use this fact to define the Levi-Civita connection and the volume form of the model spaces. We have already observed that for all non degenerate model spaces $\mathbb{M} = \mathcal{M}/\{\pm \mathrm{Id}\}$ which we have defined as the (projectivization of the) subset $\mathcal{M}$ of $(\mathbb{R}^4, b)$ where the quadratic form defined by b takes the value 1 (or -1), the normal vector to $\mathcal{M}$ at a point x is precisely x itself. Thus, by the definition of the Levi-Civita connection of an embedded hypersurface in a higher-dimensional manifold, we obtain, for vector fields v, w tangent to $\mathcal{M}$,

$$Dw(v) = \nabla^{\mathcal{M}}_v w - b(v, w)x.$$

We used that the identity and the second fundamental form coincide (up to a sign) with the first fundamental form (see also Section 6 below). Clearly this definition descends to the definition of the Levi-Civita connection of $\mathbb{M}$.

Also the volume form of $\mathcal{M}$ (and thus of $\mathbb{M}$) can be defined in terms of the ambient volume form of $\mathbb{R}^4$. Indeed, given any triple of vectors v, w, u in $T_x\mathcal{M}$, x is also the normal vector to $\mathcal{M}$, and thus one can define

$$\omega_{\mathcal{M}}(v, w, u) = \omega_{\mathbb{R}^4}(x, v, w, u).$$

5.2 Degenerate cases

We will now introduce a natural connection and volume form on ${}^*\mathbb{E}^3$ and ${}^*\mathbb{M}\mathrm{in}^3$. We will discuss the meaning of naturality in the following, but of course we can anticipate that a natural connection/volume form will be preserved by the isometry group of co-Euclidean (resp. co-Minkowski) space, as introduced in Definitions 3.10 and 3.26.

The connection of co-Euclidean space. To define a connection on ${}^*\mathbb{E}^3$, we start by defining a connection on its double cover, namely

$$co\mathcal{E}uc^3 = \{(x_1, x_2, x_3, x_4) \mid x_1^2 + x_2^2 + x_3^2 = 1\} = \{x \in \mathbb{R}^4 \mid b^*(x, x) = 1\},$$

and thus we will consider $\mathcal{S}^2 \times \mathbb{R}$ as a model for $co\mathcal{E}uc^3$. In the classical case of model spaces defined by a non degenerate symmetric quadratic form b, the key ingredient to define the Levi-Civita connection was the existence of the normal vector field N, so as to be able to write

$$Dw(v) = \nabla_v^{\mathcal{M}} w + b(v, w)\mathrm{N}.$$

Clearly the normal vector field has the property that it is preserved by the group of isometries of the ambient quadratic form b. In this degenerate case, the bilinear form b^* is degenerate, and thus it does not enable us to determine a unit normal vector field. However, there is a well-defined transverse vector field to $co\mathcal{E}uc^3$, namely the vector field which at the point $x \in co\mathcal{E}uc^3 \subset \mathbb{R}^4$ is defined by

$$\mathrm{N}_x = x \in \mathbb{R}^4,$$

Tautologically, this vector field N is preserved by the group $\mathrm{Isom}({}^*\mathbb{E}^3)$, which means that if $A \in \mathrm{Isom}({}^*\mathbb{E}^3)$, then $A_*(\mathrm{N}_x) = \mathrm{N}_{A(x)}$. Thus one can use the vector field N to decompose the ambient derivative of two vector fields in a tangential and a "normal" component.

Definition 5.1 (co-Euclidean connection). Given two vector fields v, w in $co\mathcal{E}uc^3$, we define the connection $\nabla^{co\mathcal{E}uc^3}$ by means of

$$Dw(v) = \nabla_v^{co\mathcal{E}uc^3} w + b^*(v, w)\mathrm{N},$$

The co-Euclidean connection is the connection $\nabla^{{}^*\mathbb{E}^3}$ induced on ${}^*\mathbb{E}^3$ by $\nabla^{co\mathcal{E}uc^3}$.

First, it should be clear that the co-Euclidean connection is preserved by the group of isometries of co-Euclidean space. We prove it now, and in fact this also follows from the characterization given in Proposition 5.3 below.

Lemma 5.2. *The co-Euclidean connection* $\nabla^{{}^*\mathbb{E}^3}$ *is invariant for* $\mathrm{Isom}({}^*\mathbb{E}^3)$.

Proof. From the definition, we have

$$\begin{aligned}\nabla_{A_*v}^{co\mathcal{E}uc^3} A_*w &= D(A_*w)(A_*v) - b^*(A_*v, A_*w)\mathrm{N}_{A(x)}\\ &= A_*(Dw(v) - b^*(v, w)\mathrm{N}_x)\\ &= \nabla_v^{co\mathcal{E}uc^3} w,\end{aligned}$$

since the ambient connection D and the vector field N are invariant for the action of $\mathrm{Isom}({}^*\mathbb{E}^3)$. □

We will denote by T the vector field on $co\mathcal{E}uc^3$ defined by $(0,1)$ in $T_x co\mathcal{E}uc^3 \cong T_x\mathbb{S}^2 \times \mathbb{R}$. It is a degenerate vector field invariant for the group $\mathrm{Isom}(co\mathcal{E}uc^3)$, as seen from the form inside the brackets of (15). Observe that T does not descend to a global vector field on ${}^*\mathbb{E}^3$, but we will still talk about the vector field T (by an abuse of notation) as the vector field induced on any simply connected open subset of ${}^*\mathbb{E}^3$.

Proposition 5.3. *The connection $\nabla^{{}^*\mathbb{E}^3}$ is the unique connection on ${}^*\mathbb{E}^3$ such that*

- $\nabla^{{}^*\mathbb{E}^3}$ *is* ***symmetric****, i.e.*

  $$\nabla^{{}^*\mathbb{E}^3}_X Y - \nabla^{{}^*\mathbb{E}^3}_Y X = [X,Y]$$

  for any pair of vector fields X, Y;
- $\nabla^{{}^*\mathbb{E}^3}$ *is* ***compatible*** *with the degenerate metric g^* of ${}^*\mathbb{E}^3$, i.e.*

  $$Z.g^*(X,Y) = g^*(\nabla^{{}^*\mathbb{E}^3}_Z X, Y) + g^*(X, \nabla^{{}^*\mathbb{E}^3}_Z Y)$$

  for any vector fields X, Y, Z;
- $\nabla^{{}^*\mathbb{E}^3}$ ***preserves every space-like plane of*** ${}^*\mathbb{E}^3$*, i.e. for V, W vector fields on a space-like plane P, $\nabla^{{}^*\mathbb{E}^3}_V W$ is tangent to P;*
- *the vector field* T *is* ***parallel*** *with respect to $\nabla^{{}^*\mathbb{E}^3}$, i.e. $\nabla^{{}^*\mathbb{E}^3}\mathrm{T} = 0$.*

In particular, the restriction of $\nabla^{{}^*\mathbb{E}^3}$ to any space-like plane coincides with the Levi-Civita connection for the induced metric.

Proof. It suffices to prove the statement for the double cover $co\mathcal{E}uc^3$, since clearly $-\mathrm{Id}$ acts on $co\mathcal{E}uc^3$ as an isometry of $co\mathcal{E}uc^3$, and preserving $\nabla^{co\mathcal{E}uc^3}$. It is straightforward to check that $\nabla^{co\mathcal{E}uc^3}$ defines a connection on $co\mathcal{E}uc^3$. Symmetry follows from the observation that $\nabla^{co\mathcal{E}uc^3}_X Y - \nabla^{co\mathcal{E}uc^3}_Y X$ is the tangential component of $(D_X Y - D_Y X) = [X,Y]$, which equals $[X,Y]$ itself. Also compatibility is very simple: for every vector Z tangent to $co\mathcal{E}uc^3$,

$$\begin{aligned} Z.g^*(X,Y) &= g^*(D_Z X, Y) + g^*(X, D_Z Y) \\ &= g^*(\nabla^{co\mathcal{E}uc^3}_Z X, Y) + g^*(X, \nabla^{co\mathcal{E}uc^3}_Z Y). \end{aligned}$$

For the third point, let P be a plane of $co\mathcal{E}uc^3$ obtained as the intersection of $co\mathcal{E}uc^3$ with a linear hyperplane P' of $\mathbb{R}^4$. Given vector fields V, W on P, $D_V W$ is tangent to P' and thus the projection to $co\mathcal{E}uc^3$ is still in P. Finally, it is clear from the construction that the derivative of T in any direction vanishes.

Let us now assume that the four conditions hold. In the coordinate system provided by $\mathbb{S}^2 \times \mathbb{R}$, the restriction of $\nabla^{co\mathcal{E}uc^3}$ to every plane $\mathbb{S}^2 \times \{*\}$ preserves the plane itself (by the third point) and coincides with the Levi-Civita connection of $\mathbb{S}^2$, by the second point.

Hence it is easily seen that the Christoffel symbols Γ_{ij}^k are those of the Levi-Civita connection when i, j, k correspond to coordinates of $\mathbb{S}^2$. Otherwise, using the first and fourth hypothesis, the Γ_{ij}^k vanish. Hence the connection is uniquely determined. ☐

Corollary 5.4. *The geodesics for the co-Euclidean connection* $\nabla^{{}^*\mathbb{E}^3}$ *coincide with the lines of* ${}^*\mathbb{E}^3$.

Proof. Again, we prove the statement for the double cover $co\mathcal{E}uc^3$. Given a space-like line l of $co\mathcal{E}uc^3$, using the action of $\mathrm{Isom}(co\mathcal{E}uc^3)$ we can assume that l is contained in the slice $\mathbb{S}^2 \times \{0\}$. Since the connection on such a slice coincides with the Levi-Civita connection, and lines of $co\mathcal{E}uc^3$ are geodesics for this copy of $\mathbb{S}^2$, l is a geodesic for $\nabla^{co\mathcal{E}uc^3}$. If l is not space-like, then it is of the form $\{*\} \times \mathbb{R}$. Since $\nabla^{co\mathcal{E}uc^3}_{\mathrm{T}}\mathrm{T} = 0$, it is clear by construction that l is geodesic, provided it is parametrized in such a way that its tangent vector is a fixed multiple of T for all time.

Since there is a line of $co\mathcal{E}uc^3$ through every point of $co\mathcal{E}uc^3$ with every initial velocity, this shows that all geodesics for the connection $\nabla^{co\mathcal{E}uc^3}$ are lines of $co\mathcal{E}uc^3$. ☐

The volume form of co-Euclidean space. By means of the transverse vector field N in $\mathbb{R}^4$, we can also perform the usual construction to define a volume form for co-Euclidean space. Indeed, we can give the following definition:

Definition 5.5 (co-Euclidean volume form). The volume form of $co\mathcal{E}uc^3$ is the 3-form $\omega_{co\mathcal{E}uc^3}$ such that, given vectors v, w, u in $T_x co\mathcal{E}uc^3$,

$$\omega_{co\mathcal{E}uc^3}(v, w, u) = \omega_{\mathbb{R}^4}(\mathrm{N}, v, w, u).$$

The co-Euclidean volume form is the volume form $\omega_{{}^*\mathbb{E}^3}$ induced on ${}^*\mathbb{E}^3$ by $\omega_{co\mathcal{E}uc^3}$.

Both the volume form of the ambient $\mathbb{R}^4$ and the vector field N are invariant for $\mathrm{Isom}({}^*\mathbb{E}^3)$, hence clearly:

Lemma 5.6. *The volume form* $\omega_{co\mathcal{E}uc^3}$ *is invariant for the group* $\mathrm{Isom}({}^*\mathbb{E}^3)$.

Of course, as there is no Riemannian or pseudo-Riemannian metric invariant by the group of isometries of ${}^*\mathbb{E}^3$ (Fact 3.11), there is no volume arising as the volume associated to a metric. However, the volume form $\omega_{{}^*\mathbb{E}^3}$ has the property that

$\omega_{^*\mathbb{E}^3}(v, w, \mathrm{T}) = 1$ provided v, w are orthonormal space-like vectors for the degenerate metric of $^*\mathbb{E}^3$, and T is the "unitary" degenerate vector field, so that the triple (v, w, T) is positively oriented. As in the classical case, since parallel transport preserves both the degenerate metric (this follows from the compatibility with the metric) and the vector field T (see Proposition 5.3), the volume form $\omega_{^*\mathbb{E}^3}$ has the following characterization:

Proposition 5.7. *The volume form $\omega_{^*\mathbb{E}^3}$ is the unique volume form on $^*\mathbb{E}^3$ such that*

- $\omega_{^*\mathbb{E}^3}$ *is* ***parallel*** *with respect to $\nabla^{^*\mathbb{E}^3}$, i.e. $\nabla^{^*\mathbb{E}^3}\omega_{^*\mathbb{E}^3} = 0$;*
- $\omega_{^*\mathbb{E}^3}(v, w, \mathrm{T}) = 1$*, where v, w, T is an oriented triple at a fixed point x_0, such that v, w are orthonormal space-like vectors.*

The case of co-Minkowski space. Very similar constructions can be used to define a connection and a volume form for co-Minkowski space. Indeed, we define the vector field N, which is transverse to $co\mathcal{M}in^3 \cong \mathcal{H}^2 \times \mathbb{R}$ in the ambient space $\mathbb{R}^4$, as the vector field $\mathrm{N}_x = x \in \mathbb{R}^4$. Clearly this definition is invariant by the isometries of $co\mathcal{M}in^3$, hence in particular by the involution which identifies the two connected components of $\mathcal{H}^2 \times \mathbb{R}$. For simplicity, we identify $^*\mathbb{Min}^3 \cong \mathbb{H}^2 \times \mathbb{R}$ as one of the two connected components of $\mathcal{H}^2 \times \mathbb{R}$. Recalling that the bilinear form for co-Minkowski space b^* has the form $x_1^2 + x_2^2 - x_3^2$ in the (x_1, x_2, x_3, x_4)-coordinates, using the ambient connection and volume form of $\mathbb{R}^4$, one obtains the definitions of co-Minkowski connection and volume form:

Definition 5.8 (co-Minkowski connection). Given two vector fields v, w in $^*\mathbb{Min}^3$, we define the connection $\nabla^{^*\mathbb{Min}^3}$ by means of

$$Dw(v) = \nabla_v^{^*\mathbb{Min}^3} w + b^*(v, w)\mathrm{N}.$$

Definition 5.9 (co-Minkowski volume form). The volume form of $^*\mathbb{Min}^3$ is the 3-form $\omega_{^*\mathbb{Min}^3}$ such that, given vectors v, w, u in $T_x^*\mathbb{Min}^3$,

$$\omega_{^*\mathbb{Min}^3}(v, w, u) = \omega_{\mathbb{R}^4}(\mathrm{N}, v, w, u).$$

We report on the key properties below, without giving the proofs, as they are completely analogous to the co-Euclidean case.

Lemma 5.10. *The co-Minkowski connection $\nabla^{^*\mathbb{Min}^3}$ and volume form $\omega_{co\mathcal{E}uc^3}$ are invariant under the action of the isometry group* $\mathrm{Isom}(^*\mathbb{Min}^3)$.

Here we denote by T the normalized degenerate vector field (as we recall that there is a notion of length of the degenerate direction which is preserved by the group of isometries of ${}^*\mathbb{Min}^3$). In the (x_1, x_2, x_3, x_4)-coordinates, T can be written as $(1, 0, 0, 0)$.

Proposition 5.11. *The connection* $\nabla^{{}^*\mathbb{Min}^3}$ *is the unique connection on* ${}^*\mathbb{Min}^3$ *such that*

- $\nabla^{{}^*\mathbb{Min}^3}$ *is* ***symmetric,*** *i.e.*
$$\nabla_X^{{}^*\mathbb{Min}^3} Y - \nabla_Y^{{}^*\mathbb{Min}^3} X = [X, Y]$$
for any pair of vector fields X, Y;
- $\nabla^{{}^*\mathbb{Min}^3}$ *is* ***compatible*** *with the degenerate metric* g^* *of* ${}^*\mathbb{Min}^3$, *i.e.*
$$\nabla^{{}^*\mathbb{Min}^3} g^* = 0;$$
- $\nabla^{{}^*\mathbb{Min}^3}$ ***preserves every space-like plane of*** ${}^*\mathbb{Min}^3$, *i.e. for* V, W *vector fields on a space-like plane* P, $\nabla_V^{{}^*\mathbb{Min}^3} W$ *is tangent to* P;
- *the vector field* T *is* ***parallel*** *with respect to* $\nabla^{{}^*\mathbb{Min}^3}$, *i.e.* $\nabla^{{}^*\mathbb{Min}^3}\mathrm{T} = 0$.

Corollary 5.12. *Geodesics for the co-Minkowski connection* $\nabla^{{}^*\mathbb{Min}^3}$ *coincide with lines of* ${}^*\mathbb{Min}^3$.

Proposition 5.13. *The volume form* $\omega_{{}^*\mathbb{Min}^3}$ *is the unique volume form on* ${}^*\mathbb{Min}^3$ *such that*

- $\omega_{{}^*\mathbb{Min}^3}$ *is* ***parallel*** *with respect to* $\nabla^{{}^*\mathbb{Min}^3}$, *i.e.* $\nabla^{{}^*\mathbb{Min}^3}\omega_{{}^*\mathbb{Min}^3} = 0$;
- $\omega_{{}^*\mathbb{Min}^3}(v, w, \mathrm{T}) = 1$, *where* v, w, T *is an oriented triple at a fixed point* x_0, *such that* v, w *are orthonormal space-like vectors.*

Relation with geometric transition. We will now show that the connections of co-Euclidean and co-Minkowski space are natural also in the sense of geometric transition, that is, they are the limits of the Levi-Civita connections under the transitions described in Section 4.1 and 4.2. We will focus on dimension 3 for definiteness, and consider the transition of $\mathbb{Ell}^3$ and $\mathrm{d}\mathbb{S}^3$ to ${}^*\mathbb{E}^3$, as described in Scheme (24), which makes use of the projective transformations

$$g_t^* = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \frac{1}{t} \end{bmatrix} \in \mathrm{PGL}(4, \mathbb{R}). \tag{28}$$

Proposition 5.14. *Let* X_t, Y_t, Z_t *be smooth families of smooth vector fields on* $\mathbb{E}\mathbb{l}\mathbb{l}^3$ *(resp.* $\mathrm{d}\mathbb{S}^3$*) such that* X_0, Y_0, Z_0 *are tangent to a plane* P*. The limit of* $g_t^* X_t$ *is a vector field on* ${}^*\mathbb{E}^3$ *which we denote by* $\dot{X}$*, and analogously for* $\dot{Y}$ *and* $\dot{Z}$*. Then,*

- *the limit of* $g_t^* \nabla_X^{\mathbb{E}\mathbb{l}\mathbb{l}^3} Y$ *(resp.* $g_t^* \nabla_X^{\mathrm{d}\mathbb{S}^3} Y$*) is* $\nabla_{\dot{X}}^{{}^*\mathbb{E}^3} \dot{Y}$*;*
- *the limit of* $\omega^{\mathbb{E}\mathbb{l}\mathbb{l}^3}(X_t, Y_t, Z_t)$ *(resp.* $\omega^{\mathrm{d}\mathbb{S}^3}(X_t, Y_t, Z_t)$*) is* $\omega^{{}^*\mathbb{E}^3}(\dot{X}, \dot{Y}, \dot{Z})$*.*

Proof. First, let us assume that P is defined by $x_4 = 0$ in $\mathbb{R}\mathrm{P}^3$, as in Section 4.2, so that we can use the projective transformations in (28). We can as usual perform the computation in the double cover $co\mathcal{E}uc^3$. Hence in those coordinates, the vector field X_0, Y_0, Z_0 have x_4-coordinate equal to 0. Hence, the limit of $g_t^* X_t$ is

$$\lim_{t \to 0} g_t^* X_t(x) = \lim_{t \to 0} \begin{bmatrix} (X_t(x))_1 \\ (X_t(x))_2 \\ (X_t(x))_3 \\ \dfrac{(X_t(x))_4}{t} \end{bmatrix} = \begin{bmatrix} (X_0(x))_1 \\ (X_0(x))_2 \\ (X_0(x))_3 \\ (\dot{X}(x))_4 \end{bmatrix},$$

namely, it is a vector field on ${}^*\mathbb{E}^3$ whose vertical projection is identified with the vector field $X_0(x)$.

Now, observe that g_t^* maps $\mathcal{S}^3$ to

$$\mathcal{S}_t^3 := g_t^*(\mathcal{S}^3) = \{x \mid x_1^2 + x_2^2 + x_3^2 + t^2 x_4^2 = -1\}.$$

Moreover, g_t^* is an isometry if we endow $\mathcal{S}_t^3$ with the metric induced by the quadratic form

$$b_{4,0}^t(x, x) = x_1^2 + x_2^2 + x_3^2 + t^2 x_4^2.$$

Hence g_t^* maps the normal vector $\mathrm{N}(x) = x$ of $\mathcal{S}^3$ to $\mathrm{N}_t(x) = g_t^*(x)$, which is precisely the normal vector to $\mathcal{S}_t^3$ for $b_{4,0}^t$.

Now, as X_0, Y_0, Z_0 are tangent to P, their base point x_t are rescaled to $g_t^* x_t$ which converge to x_∞, base point for $\dot{X}, \dot{Y}, \dot{Z}$. For the same reason, the normal vector $\mathrm{N}_t(x_t) = g_t^*(x_t)$ to $\mathcal{S}_t^3$ at x_t converges to x_∞, which is (in the notation of Definitions 5.1 and 5.8) the vector $\mathrm{N}(x_\infty)$ used as a transverse vector field to $co\mathcal{E}uc^3$.

Hence in conclusion,

$$g_t^* \nabla_{X_t}^{\mathcal{S}^3} Y_t = \nabla_{g_t^* X_t}^{\mathcal{S}_t^3} g_t^* Y_t = DY_t(X_t) - b_{4,0}^t(X_t, Y_t)\mathrm{N}_t(x_t)$$

converges to

$$\nabla_{\dot{X}}^{co\mathcal{E}uc^3} \dot{Y} = D\dot{Y}(\dot{X}) - b^*(\dot{X}, \dot{Y})\mathrm{N}(x_\infty).$$

Analogously,

$$\omega_{\mathbb{S}^2}(X_t, Y_t, Z_t) = \omega_{\mathbb{R}^4}(\mathrm{N}, X_t, Y_t, Z_t) = \omega_{\mathbb{E}^4}(\mathrm{N}_t, g_t^* X_t, g_t^* Y_t, g_t^* Z_t)$$

converges to

$$\omega_{co\mathcal{E}uc^3}(\dot{X}, \dot{Y}, \dot{Z}) = \omega_{\mathbb{R}^4}(\mathrm{N}(x_\infty), \dot{X}, \dot{Y}, \dot{Z}).$$

Of course, the proof is completely analogous for the rescaling from $\mathrm{d}\mathbb{S}^3$. □

Actually, the proof is similar for the cases of geometric transitions of $\mathbb{H}^3$ and $\mathbb{A}\mathrm{d}\mathbb{S}^3$ to ${}^*\mathbb{M}\mathrm{im}^3$, hence one can prove:

Proposition 5.15. *Let X_t, Y_t, Z_t be smooth families of smooth vector fields on $\mathbb{H}^3$ (resp. $\mathbb{A}\mathrm{d}\mathbb{S}^3$) such that X_0, Y_0, Z_0 are tangent to a plane P. The limit of $g_t^* X_t$ is a vector field on ${}^*\mathbb{M}\mathrm{im}^3$ which we denote by $\dot{X}$, and analogously for $\dot{Y}$ and $\dot{Z}$. Then,*

- *the limit of $g_t^* \nabla^{\mathbb{H}^3}_X Y$ (resp. $g_t^* \nabla^{\mathbb{A}\mathrm{d}\mathbb{S}^3}_X Y$) is $\nabla^{{}^*\mathbb{M}\mathrm{im}^3}_{\dot{X}} \dot{Y}$;*
- *the limit of $\omega^{\mathbb{H}^3}(X_t, Y_t, Z_t)$ (resp. $\omega^{\mathbb{A}\mathrm{d}\mathbb{S}^3}(X_t, Y_t, Z_t)$) is $\omega^{{}^*\mathbb{M}\mathrm{im}^3}(\dot{X}, \dot{Y}, \dot{Z})$.*

5.3 The infinitesimal Pogorelov map

Weyl formula for connections. In an affine chart, a (non-degenerate) model space $\mathbb{M}$ has the strong property that its (unparametrized) geodesics are the same as in the ambient $\mathbb{R}^n$ endowed with a Euclidean metric. This will imply that, in an open set, the Levi-Civita connection ∇ of $\mathbb{M}$ can be written in terms of the usual connection on $\mathbb{R}^n$. The manifold $\mathbb{M}$ may not be orientable, but on an affine chart we can consider the same orientation as $\mathbb{R}^n$. Hence the pseudo-Riemannian metric of $\mathbb{M}$ gives a volume form ω in the affine chart.

Let N be an orientable manifold with two torsion-free connections ∇ and $\widehat{\nabla}$ and with two volume forms ω and $\hat{\omega}$ such that $\nabla\omega = 0$ and $\widehat{\nabla}\hat{\omega} = 0$. Let λ be the function on N such that

$$\hat{\omega} = \lambda\omega.$$

Let us introduce

$$D(X, Y) = \widehat{\nabla}_X Y - \nabla_X Y.$$

This function is linear over smooth functions on both argument, so it is a $(1, 2)$-tensor. As the connections are symmetric, it is easy to check that D is symmetric.

Lemma 5.16. *The contraction of D is a 1-form that is given by λ:*

$$C_1^1 D = \mathrm{d}\ \ln \lambda. \tag{29}$$

Proof. Recall that the volume form ω is parallel for ∇ (26), that gives, in the given local coordinates on N,

$$\partial_i \omega_{1,\dots,n} = \omega\Big(\sum_{k=1}^{n} \Gamma_{i1}^k e_k, \dots, e_n\Big) + \dots + \omega\Big(e_1, \dots, \sum_{k=1}^{n} \Gamma_{in}^k e_k\Big).$$

But ω is alternating and n-linear:

$$\partial_i \omega_{1,\dots,n} = \omega(\Gamma_{i1}^1 e_1, \dots, e_n) + \dots + \omega(e_1, \dots, \Gamma_{in}^n e_n) = \Big(\sum_{\alpha=1}^{n} \Gamma_{i\alpha}^{\alpha}\Big)\omega_{1,\dots,n},$$

namely

$$\sum_{\alpha=1}^{n} \Gamma_{i\alpha}^{\alpha} = \partial_i \ln(\omega_{1,\dots,n})$$

and in particular,

$$\sum_{\alpha=1}^{n} \hat{\Gamma}_{i\alpha}^{\alpha} - \sum_{\alpha=1}^{n} \Gamma_{i\alpha}^{\alpha} = \partial_i \ln \lambda. \qquad \square$$

Lemma 5.17 (Weyl Formula). *Let us suppose that both ∇ and $\hat{\nabla}$ have the same (unparameterized) geodesics. Then*

$$\hat{\nabla}_X Y - \nabla_X Y = X.(\ln \lambda^{\frac{1}{n+1}})Y + Y.(\ln \lambda^{\frac{1}{n+1}})X.$$

Proof. Observe that in this formula both sides are linear in X and Y. To compute the left-hand side, fix p and $X_p \in T_p N$ and let $c\colon I \to N$ be a parameterized geodesic for ∇ with tangent vector X_p at $c(0) = p$. Extend X_p to X, the tangent vector of c, which is parallel along c, so that $\nabla_X X = 0$. Let $\hat{X}$ be the tangent vector of a reparametrization of c that turns it into a geodesic $\hat{c}$ for $\hat{\nabla}$ (hence $\hat{\nabla}_{\hat{X}} \hat{X} = 0$). Then there is a function f on c (which depends on the choice of X_p and $\hat{X}_p$) such that $X = f\hat{X}$. We compute

$$\begin{aligned}
(\hat{\nabla}_X X - \nabla_X X)(p) &= (\hat{\nabla}_X X)(p) \\
&= \hat{\nabla}_{f(p)\hat{X}_p}(f\hat{X}) \\
&= f(p)(\hat{X}_p.f)\hat{X}_p + f(p)^2 \hat{\nabla}_{\hat{X}_p} \hat{X}_p \\
&= f(p)(\hat{X}_p.f)\hat{X}_p,
\end{aligned}$$

so $D(X, X) = X.\ln(f)X$. Let us define $\phi(X) = X.\ln(f_X)/2$, so that

$$D(X, X) = 2\phi(X)X. \tag{30}$$

Let us show that ϕ is a 1-form. In fact, given any symmetric tensor D such that $D(X, X)$ is a multiple of X, the function ϕ satisfying (30) is uniquely determined and clearly satisfies $\phi(\lambda X) = \lambda\phi(X)$. For the additivity, from the symmetry of D, we have

$$D(X + Y, X + Y) = D(X, X) + D(Y, Y) + 2D(X, Y), \tag{31}$$

hence

$$\phi(X + Y)(X + Y) = \phi(X)X + \phi(Y)Y + D(X, Y).$$

On the other hand, developing the same expression for $D(X - Y, X - Y)$ gives

$$\phi(X - Y)(X - Y) = \phi(X)X + \phi(Y)Y - D(X, Y).$$

Putting together the two expressions, one obtains

$$(\phi(X + Y) + \phi(X - Y) - 2\phi(X))X = (\phi(X - Y) - \phi(X + Y) + 2\phi(Y))Y.$$

Since it now suffices to consider X and Y linearly independent, one has

$$\phi(X + Y) + \phi(X - Y) - 2\phi(X) = \phi(X - Y) - \phi(X + Y) + 2\phi(Y) = 0$$

and therefore $\phi(X + Y) = \phi(X) + \phi(Y)$. Hence, using the linearity of ϕ which was just proved, from (30) and (31) one obtains

$$D(X, Y) = \phi(X)Y + \phi(Y)X.$$

Finally, contracting on both side using (29) leads to $\mathrm{d}\ln\lambda = (n + 1)\phi$. □

Killing fields. If ∇ is the Levi-Civita connection of a pseudo-Riemannian metric g on a manifold N, a vector field is called a *Killing field* if it generates a 1-parameter group of isometries. A vector field K is a Killing field if and only if $\mathcal{L}_X g = 0$ ($\mathcal{L}$ is the Lie derivative), or if and only if ∇K is a skew-symmetric $(1, 1)$-tensor:

$$g(\nabla_X K, Y) + g(X, \nabla_Y K) = 0, \tag{32}$$

for every X, Y. It is easy to see that this condition is actually equivalent to the condition that $g(\nabla_X K, X) = 0$ for every X.

The infinitesimal Pogorelov map. Let us suppose that $\widehat{\nabla}$ is also a Levi-Civita connection for a pseudo-Riemannian metric $\hat{g}$ on N. Let $L := L_{g,\hat{g}}$ be the map $TN \to TN$ defined by

$$g(X,Y) = \hat{g}(L(X),Y) \tag{33}$$

for every Y. The *infinitesimal Pogorelov map*

$$\mathrm{P} := \mathrm{P}_{g,\hat{g}} \colon TN \longrightarrow TN$$

is defined by

$$\mathrm{P}(X) = \lambda^{\frac{2}{n+1}} L(X).$$

Lemma 5.18. *The infinitesimal Pogorelov map* $\mathrm{P}_{g,\hat{g}}$ *sends Killing fields of* g *to Killing fields of* $\hat{g}$*: if* K *is a Killing field of* (N,g)*, then* $\mathrm{P}(K)$ *is a Killing field of* $(N,\hat{g})$.

Proof. By definition,

$$\hat{g}(\mathrm{P}(K),X) = g(\lambda^{\frac{2}{n+1}}K,X), \tag{34}$$

which implies

$$X.\hat{g}(\mathrm{P}(K),X) = X.g(\lambda^{\frac{2}{n+1}}K,X),$$

and using the fact that K is a Killing field of (N,g), we arrive at

$$\hat{g}(\widehat{\nabla}_X \mathrm{P}(K),X) = (X.\lambda^{\frac{2}{n+1}})g(K,X) - g(\lambda^{\frac{2}{n+1}}K,\widehat{\nabla}_X X - \nabla_X X). \tag{35}$$

Observe that by Weyl formula (Lemma 5.17),

$$g(\lambda^{\frac{2}{n+1}}K,\widehat{\nabla}_X X - \nabla_X X) = g(\lambda^{\frac{2}{n+1}}K, 2X.(\ln \lambda^{\frac{1}{n+1}})X)) = (X.\lambda^{\frac{2}{n+1}})g(K,X).$$

Hence from equation (35) one gets $\hat{g}(\widehat{\nabla}_X \mathrm{P}(K),X) = 0$, i.e. $\mathrm{P}(K)$ is a Killing field of $(M,\hat{g})$. ☐

The infinitesimal Pogorelov map from hyperbolic to Euclidean space. Let us consider the usual Klein model of the hyperbolic space $\mathbb{H}^n$, i.e. the affine chart $\{x_{n+1} = 1\}$, in which $\mathbb{H}^n$ is the open unit ball

$$B^n = \{x \in \mathbb{R}^n \mid b_{n,0} < 1\}.$$

We will compare the hyperbolic metric g on B^n with the standard Euclidean metric. For $x \in B^n$, the point $r(x) := \rho(x)\binom{x}{1}$, with

$$\rho(x) = \frac{1}{\sqrt{1 - b_{n,0}(x,x)}}$$

belongs to $\mathcal{H}^n$ in $\mathbb{M}\mathrm{i}\mathbb{m}^{n+1}$.[17] So if $X \in T_x B^n$, then

$$\mathrm{d}\,r(X) = \rho(x)\begin{pmatrix} \rho(x)^2 b_{n,0}(x, X)x + X \\ \rho(x)^2 b_{n,0}(x, X) \end{pmatrix}$$

belongs to $T_{r(x)}\mathcal{H}^n$, and it is then straightforward that the expression of the hyperbolic metric in the Klein model is

$$g_x(X, Y) = \rho(x)^2 b_{n,0}(X, Y) + \rho(x)^4 b_{n,0}(x, X) b_{n,0}(x, Y). \tag{36}$$

The radial direction at x is the direction defined by the origin of $\mathbb{R}^n$ and x. From (36), a vector is orthogonal to the radial direction for the Euclidean metric if and only if it is orthogonal for the hyperbolic metric. A vector orthogonal to the radial direction is called lateral. Then for a tangent vector X of B^n at a point x,

- if X is radial, then its hyperbolic norm is $\rho(x)^2$ times its Euclidean norm.
- if X is lateral, then its hyperbolic norm is $\rho(x)$ times its Euclidean norm.

With respect to the definition of $L = L_{g,b_{n,0}}$ in (33),

$$L_x(X) = \rho(x)^2 X + \rho(x)^4 b_{n,0}(x, X)x$$

and a lateral (resp. radial) vector is an eigenvector for L with eigenvalue $\rho(x)^2$ (resp. $\rho(x)^4$). In an orthonormal basis for $b_{n,0}$, $\det(g_x) = \det(L_x) = \rho(x)^{2(n+1)}$, so

$$\omega_{\mathbb{R}^n} = \rho^{-(n+1)} \omega_{\mathbb{H}^n}.$$

Let K be a Killing field of $\mathbb{H}^n$. Then by Lemma 5.18, $\mathrm{P}(K) = \rho^{-2} L.(K)$ is a Euclidean Killing field. More precisely,

$$\mathrm{P}(K)_x = K_x + \rho(x)^2 b_{n,0}(x, K_x)x.$$

In particular,

- if K is lateral, the Euclidean norm of $\mathrm{P}(K)$ is equal to ρ^{-1} times the hyperbolic norm of K;
- if K is radial, the Euclidean norm of $\mathrm{P}(K)$ is equal to the hyperbolic norm of K.

---

[17] If t is the hyperbolic distance from $r(0)$ to $r(x)$, then $\rho(x) = \cosh t$.

The infinitesimal Pogorelov map from anti-de Sitter to Minkowski space. Let us consider the model of anti-de Sitter space $\mathbb{A}\mathrm{d}\mathbb{S}^n$ given by the affine chart $\{x_{n+1}{=}1\}$. Recall that $\mathbb{A}\mathrm{d}\mathbb{S}^n$ is the projective quotient of

$$\mathcal{A}d\mathcal{S}^n = \{(x, x_{n+1}) \in \mathbb{R}^{n+1} \mid b_{n-1,2} = -1\}$$

and its image in the affine chart is

$$H^n = \{x \in \mathbb{R}^n \mid b_{n-1,1} < 1\}.$$

We will compare the anti-de Sitter Lorentzian metric g on H^n with the Minkowski metric on $\mathbb{R}^n$. For $x \in H^n$, the point $\rho(x)\binom{x}{1}$, with

$$\rho(x) = \frac{1}{\sqrt{1 - b_{n-1,1}(x,x)}}$$

belongs to $\mathcal{A}d\mathcal{S}^n$. A computation similar to the hyperbolic/Euclidean case gives

$$g_x(X,Y) = \rho(x)^2 b_{n-1,1}(X,Y) + \rho(x)^4 b_{n-1,1}(x,X) b_{n-1,1}(x,Y).$$

As for the Klein model of the hyperbolic space, a vector has a radial and a lateral component, and this decomposition does not depend on the metric. We also have

$$L_x(X) = \rho(x)^2 X + \rho(x)^4 b_{n-1,1}(x,X)x,$$

whose determinant in an orthonormal basis for $b_{n-1,1}$, is $\rho(x)^{2(n+1)}$, so (recall (27))

$$\omega_{\mathbb{R}^n} = \rho^{-(n+1)} \omega_{\mathbb{A}\mathrm{d}\mathbb{S}^n}.$$

Let K be a Killing field of $\mathbb{A}\mathrm{d}\mathbb{S}^n$. Then by Lemma 5.18,

$$\rho^{-2} L_{\cdot}(K) = K + \rho^2 b_{n-1,1}(\cdot, K)\cdot$$

is a Minkowski Killing field.

Comments and references

- A symmetric connection such that there exists (locally) a parallel volume form is characterized by the fact that its Ricci tensor is symmetric, [48, Proposition 3.1].

- Let ∇ be a torsion-free complete connection on a manifold N, such that a local parallel volume form exists. An *affine field* of ∇ is a vector field that generates a 1-parameter group of transformations that preserves the connection. Let R be the curvature tensor of ∇:
$$R(X,Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z.$$
Then K is an affine field if and only if (see e.g. [41])
$$\nabla_Y \nabla_X K = -R(K,Y)X.$$
Let $\widehat{\nabla}$ be another connection on N with the same properties as ∇, both having the same unparametrized geodesics. Let $\widehat{R}$ be its curvature tensor. Then a direct computation using the Weyl formula shows that
$$\widehat{R}(X,Y)Z = R(X,Y)Z + (\nabla^2 f - \mathrm{d} f^2)(X,Z)Y - (\nabla^2 f - \mathrm{d} f^2)(Y,Z)X,$$
where ∇^2 is the Hessian for ∇. See pp. 126f in [41] for the relations between infinitesimal affine transformations and Killing fields.
- Lemma 5.18 was proved in [40] and independently in [68]. See also the end of Section 6.5.
- The simplest infinitesimal Pogorelov map is the one from Minkowski space to Euclidean space: it suffices to multiply the last coordinate of the Killing field by -1. This was noted in [34]. The term infinitesimal Pogorelov map comes from the fact that it was defined as a first-order version of the so-called Pogorelov mapping, see the end of Section 6.5. See [45], [58], and [30] for some applications.

6 Geometry of surfaces in 3-dimensional spaces

The purpose of this section is the study of embedded surfaces, with particular attention to convex surfaces, in 3-dimensional model spaces. We will first review the classical theory of embeddings of surfaces in 3-manifolds, with special attention to the constant curvature cases we have introduced so far. After that, we will define analogous notions in the degenerate spaces (for instance, the second fundamental form and the shape operator), in particular co-Euclidean and co-Minkowski spaces, and show that these notions have a good behavior with respect to both the geometry of $^*\mathbb{E}^3$ and $^*\mathbb{M}\mathrm{in}^3$, although the ambient metric is a degenerate metric, and with respect to duality and geometric transition.

6.1 Surfaces in non degenerate constant curvature 3-manifolds

Let (M, g) be a Riemannian or pseudo-Riemannian manifold. Given a smooth immersion $\sigma\colon S \to M$ with image a space-like surface $\sigma(S)$ in (M, g), recall that the *first fundamental form* is the pull-back of the induced metric, namely

$$\mathrm{I}(v, w) = g(\sigma_*(v), \sigma_*(w)).$$

The Levi-Civita connection ∇^S of the first fundamental form I of S is obtained from the Levi-Civita connection of the ambient manifold: given vector fields v, w on S, $\nabla^S_v w$ is the orthogonal projection to the tangent space of S of $\nabla^M_{\sigma_* v}(\sigma_* w)$.

Let us denote by N a unit normal vector field on S, namely N is a smooth vector field such that for every point $x \in S$, N_x is orthogonal to $T_{\sigma(x)}\sigma(S)$, and such that $|g(\mathrm{N}, \mathrm{N})| = 1$. The *second fundamental form* II is a bilinear form on S defined by

$$\nabla^M_{\sigma_* v}(\sigma_* \hat{w}) = \nabla^S_v \hat{w} + \mathrm{II}(v, w)\mathrm{N} \tag{37}$$

where $\hat{w}$ denotes any smooth vector field on S extending the vector w. This form turns out to be symmetric and it only depends on the vectors v and w, not on the extension of any of them. The *shape operator* of S is the $(1, 1)$-tensor $B \in \mathrm{End}(TS)$, self-adjoint with respect to I, such that

$$\mathrm{II}(v, w) = \mathrm{I}(B(v), w) = \mathrm{I}(v, B(w)). \tag{38}$$

By a standard computation, it turns out that

$$B(v) = \pm\nabla^M_v \mathrm{N}, \tag{39}$$

the sign depending on whether M is Riemannian or Lorentzian, where we have used the differential of the embedding σ to identify a vector $v \in T_x S$ to a vector tangent to the embedded surface $\sigma(S)$. Indeed, by applying the condition of compatibility of the metric of the Levi-Civita connection to the condition $|g(\mathrm{N}, \mathrm{N})| = 1$, it is easily checked that $\nabla^M_v \mathrm{N}$ is orthogonal to N, hence is tangent to $\sigma(S)$. Since B is self-adjoint with respect to I, it is diagonalizable at every point. The eigenvalues of B are called *principal curvatures*.

The *extrinsic curvature* is the determinant of the shape operator, i.e. the product of the principal curvatures. The *mean curvature* is the trace of the shape operator, that is the sum of the principal curvatures.

Fundamental theorem of immersed surfaces. The *embedding data* of a smooth immersed surface in a 3-manifold are usually considered as the first fundamental form and the shape operator, or equivalently, the first and second fundamental forms. However, these two objects are not independent and satisfy some coupled differential equations usually called the *Gauss–Codazzi* equations. We start by expressing such equations in the setting of the constant curvature 3-manifolds introduced above, namely Euclidean and Minkowski space, and the model spaces.

The *Codazzi equation* can be expressed in the same fashion for all ambient spaces, and it says that the exterior derivative of the shape operator B, with respect to the Levi-Civita connection of the first fundamental form (denoted by ∇^{I}), vanishes. In formulae,

$$\mathrm{d}^{\nabla^{\mathrm{I}}} B := (\nabla^{\mathrm{I}}_v B)(w) - (\nabla^{\mathrm{I}}_w B)(v) = \nabla^{\mathrm{I}}_v B(w) - \nabla^{\mathrm{I}}_w B(v) - B[v, w] = 0. \tag{40}$$

The *Gauss equation* is a relation between the intrinsic curvature of the first fundamental form, and the extrinsic curvature of the immersion. The form of the Gauss equation, however, depends on the ambient metric (whether it is Riemannian or pseudo-Riemannian) and on its curvature. In particular, if K_{I} denotes the curvature of the first fundamental form, in Euclidean space the Gauss formula is

$$K_{\mathrm{I}} = \det B,$$

while in Minkowski space, for immersed space-like surfaces, the correct form is

$$K_{\mathrm{I}} = -\det B.$$

As we said, if the ambient manifold has nonzero curvature, there is an additional term in the equation. For instance, for $\mathbb{S}^3$ or $\mathbb{Ell}^3$ (curvature 1 everywhere),

$$K_{\mathrm{I}} = 1 + \det B,$$

while for surfaces in hyperbolic space we have

$$K_{\mathrm{I}} = -1 + \det B.$$

Finally for space-like surfaces in Lorentzian manifolds of nonzero constant curvature: in de Sitter space the Gauss equation is

$$K_{\mathrm{I}} = 1 - \det B,$$

and in anti-de Sitter space of course

$$K_{\mathrm{I}} = -1 - \det B.$$

We summarize all the statements in the following:

Fact 6.1. *Given a smooth immersed surface in a three-dimensional model space (or in Euclidean/Minkowski space), the first fundamental form and the shape operator satisfy the corresponding Gauss–Codazzi equations.*

Clearly, if one post-composes a smooth immersion with an isometry of the ambient space, the embedding data remain unchanged. A classical theorem in Euclidean space, which can be extended to all the other cases of constant curvature, states that the Gauss–Codazzi equations are necessary and sufficient conditions in order to have the embedding data of a smooth surface, and the embedding data determine the surface up to global isometry. See [7] and [63] for a reference.

Theorem 6.2 (fundamental theorem of immersed surfaces). *Given a simply connected surface S and a pair (I, B) of a Riemannian metric and a symmetric $(1,1)$-tensor on S, if (I, B) satisfy the Gauss–Codazzi equations in Euclidean (resp. Minkowski, elliptic, hyperbolic, de Sitter or anti-de Sitter) space, then there exists a smooth immersion of S having (I, B) as embedding data. Any two such immersions differ by post-composition by a global isometry.*

Duality for smooth surfaces. We can use the description of duality introduced in Section 2.4 to talk about duality for convex surfaces. In fact, consider a surface S, which is the boundary of a convex set. Then its dual is again a convex set with boundary a surface S^*. For instance, the dual of the boundary of a convex set in $\mathbb{H}^3$ is a space-like surface in $\mathrm{d}\mathbb{S}^3$ (and vice versa). Similarly one can consider a convex set in $\mathbb{A}\mathrm{d}\mathbb{S}^3$, whose boundary is composed of two space-like surfaces, and dually one obtains a convex set in $\mathbb{A}\mathrm{d}\mathbb{S}^3$ bounded by two space-like surfaces.

Let us consider a convex set with boundary a smooth (or at least C^2) embedded surface S, such that B is positive definite at every point (which in particular implies strict convexity). The *third fundamental form* of S is

$$\mathrm{III}(v, w) = \mathrm{I}(B(v), B(w)).$$

First, observe that, since B is positive definite, III is a Riemannian metric. The reader can check, as an exercise, that in his/her favorite duality of ambient spaces, (III, B^{-1}) are the embedding data of a space-like surface, namely they satisfy the Gauss–Codazzi equations provided (I, B) satisfy the Gauss–Codazzi equations. It is helpful to use the following formula for the curvature of $\mathrm{III}(v, w) = \mathrm{I}(B\cdot, B\cdot)$,

$$K_{\mathrm{III}} = \frac{K_{\mathrm{I}}}{\det B},$$

which holds under the assumption that B satisfies the Codazzi equation (see [44] or [43]).

Fact 6.3. *The pair* (III, B^{-1}) *are the embedding data of the dual surface* S^*.

Let us check this fact in the hyperbolic-de Sitter case. If S is a convex surface in $\mathbb{H}^3$, for $x \in S$, $\mathrm{N}(x)$ is a point inside the unit tangent sphere of $T_x\mathbb{H}^3$, and III is by definition the pull-back by N on S of the spherical metric. But on the other hand, the hyperplane tangent to S at x is naturally identified, via the double cover of $\mathbb{M}\mathrm{in}^4$, to a time-like hyperplane, and $\mathrm{N}(x)$ is a unit vector orthogonal to this hyperplane, hence a point in $d\mathcal{S}^n$, the double cover of de Sitter space. So III is exactly the induced metric on the dual surface, and by the involution property of the duality, its shape operator is the inverse of the one of B.

6.2 Geometry of surfaces in co-Euclidean space

We are now ready to define the second fundamental form of any space-like surface in ${}^*\mathbb{E}^3$. For simplicity (since this is a local notion) we will use the double cover $co\mathcal{E}uc^3 \cong \mathcal{S}^2 \times \mathbb{R}$. A space-like surface in $co\mathcal{E}uc^3$ is locally the graph of a function $u\colon \Omega \to \mathbb{R}$, for $\Omega \subseteq \mathcal{S}^2$, and the first fundamental form is just the spherical metric on the base $\mathcal{S}^2$.

Given a space-like immersion $\sigma\colon S \to co\mathcal{E}uc^3$, again there is no notion of normal vector field, since the metric of $co\mathcal{E}uc^3$ is degenerate. However, we have already remarked several times that the vector field T (of zero length for the degenerate metric, see Section 5.2) is well-defined, i.e. it is invariant by the action of the isometry group. Given two vector fields $\hat{v}$, $\hat{w}$ on S, using symmetry and compatibility with the metric, it is easy to prove that the tangential component of $\nabla^{co\mathcal{E}uc}_{\sigma_* v}(\sigma_*\hat{w})$ in the splitting

$$T_{(x,t)}co\mathcal{E}uc^3 = T_{(x,t)}\sigma(S) \oplus \langle \mathrm{T} \rangle,$$

coincides with the Levi-Civita connection of the first fundamental form. This enables to give the following definition:

Definition 6.4. Given a space-like immersion $\sigma\colon S \to co\mathcal{E}uc^3$, the second fundamental form of S is defined by

$$\nabla^{co\mathcal{E}uc^3}_{\sigma_* v}(\sigma_*\hat{w}) = \nabla^{\mathrm{I}}_v(\hat{w}) + \mathrm{II}(v,w)\mathrm{T},$$

for every pair of vectors $v, w \in T_xS$, where $\hat{w}$ is any extension on S of the vector $w \in T_xS$.

It is easy to check, as in the classical Riemannian case, that II is linear in both arguments, and thus defines a $(0,2)$-tensor.

Definition 6.5. The shape operator of $\sigma\colon S \to co\mathcal{E}uc^3$ is the symmetric $(1,1)$-tensor B such that $\mathrm{II}(v,w) = \mathrm{I}(B(v),w)$ for every $v,w \in T_xS$, where I is the first fundamental form of S. The extrinsic curvature of S is the determinant of the shape operator.

As usual, the definition does not depend on the extension of any of the vectors v and w and the second fundamental form is symmetric.

Lemma 6.6. *Given a space-like embedded graph S in $co\mathcal{E}uc^3$, consider the embedding $\sigma\colon \Omega \to co\mathcal{E}uc^3 \cong \mathbb{S}^2 \times \mathbb{R}$ defining S as a graph:*

$$\sigma(x) = (x, u(x))$$

for $u\colon \Omega \to \mathbb{R}$ and $\Omega \subseteq \mathbb{S}^2$. Then the shape operator of S for the embedding σ is

$$B = \mathrm{Hess}^{\mathbb{S}^2} u + u\,\mathrm{Id},$$

where $\mathrm{Hess}^{\mathbb{S}^2} u = \nabla^{\mathbb{S}^2}\operatorname{grad} u$ *denotes the spherical Hessian of u.*

Before proceeding to the proof of this proposition, let us relate functions on $\Omega \subset \mathbb{S}^2 \subset \mathbb{E}^3$ to functions on an open set of $\mathbb{E}^3$. As the unit outward vector field of $\mathbb{S}^2$ is the identity, it follows from (37) and (39) that, at a point $x \in \mathbb{S}^2$, for X, Y tangent to the sphere,

$$D_X Y = \nabla^{\mathbb{S}^2}_X Y - \langle X, Y\rangle x \tag{41}$$

where $\langle\cdot,\cdot\rangle := b_{3,0}$. From this it follows immediately that, for U the 1-homogenous extension of u,

$$\begin{aligned}\langle \mathrm{Hess}\, U(X), Y\rangle &= X.Y.U - (D_X Y).U \\ &= \mathrm{Hess}^{\mathbb{S}^2} U|_{\mathbb{S}^2}(X,Y) + \langle X, Y\rangle x.U,\end{aligned}$$

and as U is 1-homogenous, by the Euler theorem, $x.U = U$, so on $\mathbb{S}^2$

$$\mathrm{Hess}\, U = \mathrm{Hess}^{\mathbb{S}^2} u + u\,\mathrm{Id}\,. \tag{42}$$

Proof. Fix a point $x_0 \in \Omega$, and denote by $b^* = x_1y_1 + x_2y_2 + x_3y_3$ the ambient degenerate quadratic form. By composing with an element in $\mathrm{Isom}(co\mathcal{E}uc^3)$ of the form $(x,t) \mapsto (x, t + f(x))$, where $f(x) = \langle x, p_0\rangle$, we can assume S is tangent to the horizontal plane $\mathbb{S}^2 \times \{0\}$ at x_0. Indeed, observe that by the usual duality, this statement is equivalent to saying that the point in $\mathbb{E}^3$ dual to $T_{\sigma(x_0)}S$ is the origin and the plane in $\mathbb{E}^3$ dual to x_0 is $x_0^\perp$. Such a condition can be easily obtained by applying a translation of $\mathbb{E}^3$. From (42), $\operatorname{Hess}^{\mathbb{S}^2} f + f\,\mathrm{Id} = 0$, and thus it suffices to prove the statement in this case.

We consider $\mathbb{S}^2 \times \{0\}$ inside the copy of $\mathbb{E}^3$ obtained as $\{x_4 = 0\}$ in $\mathbb{R}^4$. Let $\hat{v}$, $\hat{w}$ be two vector fields on $\Omega \subseteq \mathbb{S}^2$. Then at $x_0 \in \mathbb{S}^2$ one has as in (41)

$$D_{\hat{v}}\hat{w} = \nabla^{\mathbb{S}^2}_{\hat{v}}\hat{w} - \langle \hat{v}, \hat{w}\rangle x_0.$$

Consider now the vector fields $\sigma_*(\hat{v}) = \hat{v} + \mathrm{d}u(\hat{v})\mathrm{T}$ and $\sigma_*(\hat{w}) = \hat{w} + \mathrm{d}u(\hat{w})\mathrm{T}$ on S. We choose extensions V and W in a neighbourhood of S which are invariant by translations $t \mapsto t + t_0$ in the degenerate direction of $\mathbb{S}^2 \times \mathbb{R}$. We can now compute

$$\begin{aligned} D_V W &= D_{\hat{v}}\hat{w} + D_{\hat{v}}(\mathrm{d}u(\hat{w})\mathrm{T}) + \mathrm{d}u(\hat{v})D_{\mathrm{T}}W \\ &= \nabla^{\mathbb{S}^2}_{\hat{v}}\hat{w} - \langle \hat{v}, \hat{w}\rangle x_0 + \hat{v}.\langle \operatorname{grad} u, \hat{w}\rangle \mathrm{T} \\ &= \nabla^{\mathbb{S}^2}_{\hat{v}}\hat{w} - \langle \hat{v}, \hat{w}\rangle x_0 + (\langle \operatorname{Hess}^{\mathbb{S}^2} u(\hat{v}), \hat{w}\rangle + \langle \operatorname{grad} u, \nabla^{\mathbb{S}^2}_{\hat{v}}\hat{w}\rangle)\mathrm{T}, \end{aligned}$$

where in the first equality we have substituted the expressions for V and W, and in the second equality we have used that $D\mathrm{T} = 0$ and that the chosen extension W is invariant along the direction of T. Recalling that the connection $\nabla^{co\mathcal{E}uc^3}_V W$ at x_0 is defined as the tangential component of $D_V W$ with respect to the transverse vector $\mathrm{N}_{x_0} = x_0$ and using that $\operatorname{grad} u$ vanishes at x_0 by construction, we get at x_0

$$\nabla^{co\mathcal{E}uc^3}_V W = \nabla^{\mathbb{S}^2}_{\hat{v}}\hat{w} + \mathrm{I}(\operatorname{Hess}^{\mathbb{S}^2} u(v), w).$$

By Proposition 5.3, $\nabla^{\mathbb{S}^2}_{\hat{v}}\hat{w}$ is tangent to the slice $\mathbb{S}^2 \times \{0\}$ itself, and thus the second fundamental form is

$$\mathrm{II}(v, w) = \mathrm{I}(\operatorname{Hess}^{\mathbb{S}^2} u(v), w).$$

Since $u(x_0) = 0$, this concludes the proof. □

Corollary 6.7. *A space-like surface in* $^*\mathbb{E}^3$ *is a plane if and only if* $B \equiv 0$.

Proof. Totally geodesic planes in $co\mathcal{E}uc^3$ are graphs of functions of the form $u(x) = \langle x, p\rangle$, which are functions such that $\operatorname{Hess}^{\mathbb{S}^2} u + u\,\mathrm{Id}$ vanishes. Since the statement is local, this is true also in $^*\mathbb{E}^3$. □

Recall that the induced metric on a space-like surface S in $co\mathcal{E}uc^3$, written as a graph over $\Omega \subset \mathbb{S}^2$, is just the spherical metric of $\mathbb{S}^2$. Hence the first fundamental form I of S satisfies

$$K_{\mathrm{I}} = 1. \tag{43}$$

Of course this statement is local, and thus holds in $^*\mathbb{E}^3$ as well. Moreover, it turns out that the shape operator satisfies the Codazzi equation with respect to I:

$$\mathrm{d}^{\nabla^{\mathrm{I}}} B = 0. \tag{44}$$

Indeed, let U be the 1-homogenous extension of u. Let $\widetilde{B} = \operatorname{Hess} U$. Each component $\widetilde{B}^i$, $i = 1, 2, 3$, is a 1-form on $\mathbb{R}^3$,

$$\widetilde{B}^i = \widetilde{B}^i_1 \, \mathrm{d}x^1 + \widetilde{B}^i_2 \, \mathrm{d}x^2 + \widetilde{B}^i_3 \, \mathrm{d}x^3,$$

with $U_{ij} = \widetilde{B}^i_j$. By the Schwarz lemma, $\widetilde{B}^i$ is closed, and by (40) and (41), B satisfies the Codazzi equation on $\mathbb{S}^2$. See also [35, 2.61] for the relations between closedness of 1-forms and the Codazzi equation.

The above two equations (43) and (44) can be interpreted as a *baby-version* of the Gauss–Codazzi equations for co-Euclidean geometry. Of course this is a very simple version, since the two equations are not really coupled: the first equation is independent of B. In fact, the proof of the fundamental theorem of immersions will be very simple. But before that, let us show that any tensor satisfying the Codazzi equation on $\mathbb{S}^2$ can be locally written as $\operatorname{Hess}^{\mathbb{S}^2} u + u \operatorname{Id}$.

Let Ω be a domain in $\mathbb{S}^2$ and let $\widetilde{\Omega}$ be the cone over Ω from the origin of $\mathbb{E}^3$. Note that $\widetilde{\Omega}$ is star-shaped from the origin. Let B be a $(1,1)$-tensor on Ω. Let $\widetilde{B}$ be a $(1,1)$-tensor of $\widetilde{\Omega}$ which satisfies $\widetilde{B}_x(x) = 0$, and for $\langle X, x \rangle = 0$, $\widetilde{B}_x(X) = B_{x/\|x\|}(X)$, where $\| \cdot \|$ is the usual norm on $\mathbb{E}^3$.

Suppose that B satisfies the Codazzi equation on Ω, and consider $\widetilde{B}$ as an $\mathbb{R}^3$-valued 1-form. From (40) and (41), $\widetilde{B}$ satisfies the Codazzi equation on $\widetilde{\Omega}$, hence is a closed form, and by the Poincaré Lemma, there exists $F^i \colon \mathbb{R}^3 \to \mathbb{R}$ such that $\widetilde{B}^i = \mathrm{d}F^i$. Suppose moreover that B is symmetric. This readily implies that $\widetilde{B}$ is symmetric, and hence that the 1-form $\omega = F^1 \, \mathrm{d}x^1 + F^2 \, \mathrm{d}x^2_2 + F^3 \, \mathrm{d}x^2_3$ is closed, so by the Poincaré Lemma, there exists a function $U \colon \widetilde{\Omega} \to \mathbb{R}$ with $\omega = \mathrm{d}U$, i.e. $\widetilde{B} = \operatorname{Hess} U$. From (42), we obtain the following fact, previously noted by D. Ferus in [12]:

Fact 6.8. *Let B be a tensor satisfying the Codazzi equation on a surface of constant curvature equal to* $+1$. *Then locally there exists a function u on $\mathbb{S}^2$ such that* $B = \operatorname{Hess}^{\mathbb{S}^2} u + u \operatorname{Id}$.

Proposition 6.9 (Fundamental theorem of immersed surfaces in co-Euclidean geometry). *Given a simply connected surface S and a pair* (I, B) *of a Riemannian metric and a symmetric* $(1,1)$*-tensor on S, if* (I, B) *are such that* I *has constant curvature* $+1$,

$$K_{\mathrm{I}} = 1,$$

and B is satisfies the Codazzi conditions for I,

$$\mathrm{d}^{\nabla^{\mathrm{I}}} B = 0.$$

Then there exists a smooth immersion of S in ${}^*\mathbb{E}^3$ *having* (I, B) *as embedding data. Any two such immersions differ by post-composition by a global isometry of* ${}^*\mathbb{E}^3$.

Proof. Again, it suffices to prove first the immersion statement in $co\mathcal{E}uc^3$. By Fact 6.8 there exists a function $u\colon S \to \mathbb{R}$ such that $B = \mathrm{Hess}^{\mathbb{S}^2} u + u\,\mathrm{Id}$. Let $\mathrm{dev}\colon \widetilde{S} \to \mathbb{S}^2$ be a developing map (that is, a local isometry) for the spherical metric I on S. Then we define

$$\sigma\colon S \longrightarrow co\mathcal{E}uc^3 \cong \mathbb{S}^2 \times \mathbb{R}$$

by means of

$$\sigma(x) = (\mathrm{dev}(x), u(x)).$$

Since dev is a local isometry, and the degenerate metric of $co\mathcal{E}uc^3$ reduces to the spherical metric on the $\mathbb{S}^2$ component, the first fundamental form of σ is I. By Lemma 6.6, the shape operator is B.

Given any other immersion σ' with embedding data (I, B), the projection to the first component is a local isometry, hence it differs from dev by postcomposition by a global isometry A of $\mathbb{S}^2$. By composing with an isometry of $co\mathcal{E}uc^3$ which acts on $\mathbb{S}^2$ by means of A and leaves the coordinate t invariant, we can assume the first component of σ and σ' in $\mathbb{S}^2 \times \mathbb{R}$ coincide.

Hence we have $\sigma'(x) = (\mathrm{dev}(x), u'(x))$ where u' is such that $B = \mathrm{Hess}^{\mathbb{S}^2} u' + u'\,\mathrm{Id}$. Therefore $\mathrm{Hess}^{\mathbb{S}^2}(u - u') + (u - u')\,\mathrm{Id} = 0$, which implies

$$u(x) - u'(x) = \langle \mathrm{dev}(x), p_0 \rangle$$

for some $p_0 \in \mathbb{E}^3$. This shows that σ and σ' differ by the isometry

$$(x, t) \longmapsto (x, t + \langle x, p_0 \rangle)$$

which is an element of $\mathrm{Isom}(co\mathcal{E}uc^3)$. This concludes the proof. ☐

Let us remark that (recalling the discussion of Section 3.3) if S is a space-like surface which is homotopic to $\mathbb{S}^2 \times \{\star\}$ in $co\mathcal{E}uc^3 \cong \mathbb{S}^2 \times \mathbb{R}$ and its shape operator B is positive definite, then S is the graph of u where u is a support function of a convex body whose boundary is S. Let us now compute the embedding data in Euclidean space of the dual surface of S. The formulae we will obtain are exactly the same as in the $\mathbb{A}\mathrm{d}\mathbb{S}^3$-$\mathbb{A}\mathrm{d}\mathbb{S}^3$ duality and in the $\mathbb{H}^3$-$\mathrm{d}\mathbb{S}^3$ duality.

Corollary 6.10. *Given a space-like smooth surface S in $\mathbb{E}^3$ (resp. ${}^*\mathbb{E}^3$) which bounds a convex set and with positive definite shape operator, if the embedding data of S are* (I, B)*, then the boundary of the dual convex set in ${}^*\mathbb{E}^3$ (resp. $\mathbb{E}^3$) has embedding data* (III, B^{-1}).

Proof. If the result is proved for one case, then the other case is obvious, since it suffices to exchange the roles of (I, B) and (III, B^{-1}). Let us consider a surface S in $\mathbb{E}^3$, and let S^* be the dual surface in ${}^*\mathbb{E}^3$.

Let us assume that in the double cover $co\mathcal{E}uc^3$, S^* is locally the graph of u over an open subset Ω of $\mathbb{S}^2$. To check that the first fundamental form of the dual surface is the third fundamental form, it suffices to compute the pull-back of the spherical metric of $\mathbb{S}^2$ by means of the first component of the dual embedding $\sigma\colon S \to {}^*\mathbb{E}^3$ for S^*; that is, the map which associates to a tangent plane $P = T_xS$ to S the dual point $P^* \in {}^*\mathbb{E}^3$. Again in the double cover, the first component of the dual embedding is precisely the Gauss map G of S, and the derivative of the Gauss map is the shape operator, hence one gets

$$\langle \mathrm{d}\sigma(v), \mathrm{d}\sigma(w)\rangle = \langle \mathrm{d}G(v), \mathrm{d}G(w)\rangle = \mathrm{I}(B(v), B(w)).$$

Moreover, one can directly show that the inverse of the shape operator of S, by means of the inverse of the Gauss map $G^{-1}\colon \mathbb{S}^2 \to S \subset \mathbb{E}^3$, is $B^{-1} = \mathrm{Hess}^{\mathbb{S}^2} u + u\,\mathrm{Id}$, where $u\colon \mathbb{S}^2 \to \mathbb{R}$ is the support function of the convex set bounded by S. Using Lemma 6.6, this concludes the proof. ☐

6.3 Geometry of surfaces in co-Minkowski space

By means of a very analogous construction, one can study the geometry of surfaces in co-Minkowski space. We will not repeat all the details, but just state the parallel results. Here $\langle\cdot,\cdot\rangle := b_{2,1}$ will denote the Minkowski product of $\mathbb{M}\mathrm{in}^3$, and we will implicitly identify ${}^*\mathbb{M}\mathrm{in}^3$, which is the quotient of $co\mathcal{M}in^3 \cong \mathcal{H}^2 \times \mathbb{R}$ by the action of $\{\pm\,\mathrm{Id}\}$, to one of its connected components, thus considering the model of $\mathbb{H}^2 \times \mathbb{R}$. The definition of second fundamental form of an immersion $\sigma\colon S \to {}^*\mathbb{M}\mathrm{in}^3$ makes again use of the splitting:

$$T_x co\mathcal{M}in^3 = T_x\sigma(S) \oplus \langle \mathrm{T}\rangle.$$

Definition 6.11. Given a space-like immersion $\sigma\colon S \to {}^*\mathbb{M}\mathrm{in}^3$, the second fundamental form of S is defined by

$$\nabla^{{}^*\mathbb{M}\mathrm{in}^3}_{\sigma_* v}(\sigma_*\hat{w}) = \nabla^{\mathrm{I}}_v(\hat{w}) + \mathrm{II}(v, w)\mathrm{T},$$

for every pair of vectors $v, w \in T_xS$, where $\hat{w}$ is any extension on S of the vector $w \in T_xS$.

Definition 6.12. The shape operator of $\sigma\colon S \to {}^*\mathbb{M}\mathrm{in}^3$ is the symmetric $(1,1)$-tensor such that $\mathrm{II}(v, w) = \mathrm{I}(B(v), w)$ for every $v, w \in T_xS$, where I is the first fundamental form of S. The extrinsic curvature of S is the determinant of the shape operator.

Lemma 6.13. *Given a space-like embedded graph S in* $^*\mathbb{Min}^3$*, consider the embedding* $\sigma\colon \Omega \to {}^*\mathbb{Min}^3 \cong \mathbb{H}^2 \times \mathbb{R}$ *defining S as a graph*:

$$\sigma(x) = (x, u(x))$$

for $u\colon \Omega \to \mathbb{R}$ *and* $\Omega \subseteq \mathbb{H}^2$*. Then the shape operator of S for the embedding σ is*

$$B = \mathrm{Hess}^{\mathbb{H}^2} u - u\,\mathrm{Id},$$

where $\mathrm{Hess}^{\mathbb{H}^2} u = \nabla^{\mathbb{H}^2} \operatorname{grad} u$ *denotes the hyperbolic Hessian of u.*

Corollary 6.14. *A space-like surface in* $^*\mathbb{Min}^3$ *is a plane if and only if* $B \equiv 0$.

Again, one can observe that the embedding data of a smooth space-like surface in $^*\mathbb{Min}^3$ satisfy the very simple condition that the first fundamental form I is hyperbolic, and the shape operator satisfies the Codazzi equation for I.

Proposition 6.15 (fundamental theorem of immersed surfaces in co-Minkowski geometry). *Given a simply connected surface S and a pair* (I, B) *of a Riemannian metric and a symmetric* $(1,1)$*-tensor on S, if* (I, B) *are such that* I *has constant curvature* -1*, that is,*

$$K_{\mathrm{I}} = -1,$$

and B satisfies the Codazzi conditions for I*, that is,*

$$d^{\nabla^{\mathrm{I}}} B = 0.$$

then there exists a smooth immersion of S in $^*\mathbb{Min}^3$ *having* (I, B) *as embedding data. Any two such immersions differ by post-composition by a global isometry of* $^*\mathbb{Min}^3$.

Finally, the formulae for the embedding data of dual surfaces work well in the $\mathbb{Min}^3$-$^*\mathbb{Min}^3$ duality:

Corollary 6.16. *Given a space-like smooth surface S in* $\mathbb{Min}^3$ (*resp.* $^*\mathbb{Min}^3$) *which bounds a convex set and with positive definite shape operator, if the embedding data of S are* (I, B)*, then the boundary of the dual convex set in* $^*\mathbb{Min}^3$ (*resp.* $\mathbb{Min}^3$) *has embedding data* (III, B^{-1}).

6.4 Geometric transition of surfaces

We now conclude by showing that the notion of curvature in co-Euclidean and co-Minkowski geometry, under the procedures of geometric transition we have encountered, is the rescaled limit of the usual notions in the model spaces which degenerate to $^*\mathbb{E}^3$ and $^*\mathbb{M}\mathrm{in}^3$. In particular, we will focus on the transitions defined in Section 4.2, having limit in co-Euclidean and co-Minkowski spaces.

Proposition 6.17. *Suppose σ_t is a C^2 family of space-like smooth immersions of a simply connected surface S into $\mathbb{E}\mathrm{ll}^3$ or $\mathrm{d}\mathbb{S}^3$ (resp. $\mathbb{A}\mathrm{d}\mathbb{S}^3$ or $\mathbb{H}^3$), such that σ_0 is contained in a totally geodesic plane P. Let*

$$\sigma = \lim_{t\to 0}(g_t^* \circ \sigma_t)$$

be the rescaled immersion in $^*\mathbb{E}^3$ (resp. $^*\mathbb{M}\mathrm{in}^3$) obtained by blowing-up the plane P. Then,

- *the first fundamental form of σ coincides with the first fundamental form of σ_0:*

$$\mathrm{I}(v,w) = \lim_{t\to 0} \mathrm{I}_t(v,w);$$

- *the second fundamental form of σ is the first derivative of the second fundamental form of σ_t:*

$$\mathrm{II}(v,w) = \lim_{t\to 0} \frac{\mathrm{II}_t(v,w)}{t};$$

- *the shape operator B of σ is the first derivative of the shape operator B_t of σ_t:*

$$B(v) = \lim_{t\to 0} \frac{B_t(v)}{t};$$

- *the extrinsic curvature K^{ext} of σ is the second derivative of the extrinsic curvature $K_t^{\mathrm{ext}} = \det B_t$ of σ_t:*

$$K^{\mathrm{ext}}(x) = \lim_{t\to 0} \frac{K_t^{\mathrm{ext}}(x)}{t^2}.$$

Proof. As in Section 4.2 (recall also the proof of Proposition 5.14), we can assume that the totally geodesic plane P is defined by $x_4 = 0$ in $\mathbb{R}\mathrm{P}^3$, and use the projective transformations

$$g_t^* = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \frac{1}{t} \end{bmatrix} \in \mathrm{PGL}(4,\mathbb{R}).$$

Hence, the limit of $g_t^*\sigma_t$ is

$$\lim_{t\to 0} g_t^*\sigma_t(x) = \lim_{t\to 0}\begin{bmatrix}(\sigma_t(x))_1\\(\sigma_t(x))_2\\(\sigma_t(x))_3\\\dfrac{(\sigma_t(x))_4}{t}\end{bmatrix} = \begin{bmatrix}(\sigma_0(x))_1\\(\sigma_0(x))_2\\(\sigma_0(x))_3\\\dfrac{(\dot\sigma(x))_4}{t}\end{bmatrix},$$

namely, $\sigma(x) = (\sigma_0(x), u(x))$ for some function u which encodes the derivative of the vertical component of σ_t. It is then clear, from the degenerate form of the metric of both ${}^*\mathbb{E}^3$ and ${}^*\mathbb{M}\mathbb{i}\mathbb{n}^3$, that in both cases the first fundamental form of the rescaled limit σ is the limit of the first fundamental form of σ_t.

For the second point, we will focus on the case of $\mathbb{E}\mathbb{l}\mathbb{l}^3$ (with limit in $\mathbb{E}^3$) for definiteness. Since the statement is local, we can consider the computation in the dual covers $\mathcal{S}^3$ and $co\mathcal{E}uc^3$. Consider the unit normal vector fields N_t to $\sigma_t(S)$, chosen so that at time $t = 0$ the vector field is $(0, 0, 0, 1)$ and N_t varies continuously with t. Recall that the second fundamental form of $\sigma_t(S)$ satisfies

$$\nabla^{\mathcal{S}^3}_{\sigma_t v}(\sigma_t\hat{w}) - \nabla^{\mathrm{I}_t}_v(\hat{w}) = \mathrm{II}_t(v,w)\mathrm{N}_t = \frac{\mathrm{II}_t(v,w)}{t}t\mathrm{N}_t.$$

Now applying the transformation g_t^*, we have

- $g_t^*\nabla^{\mathcal{S}^3}_{\sigma_t v}(\sigma_t\hat{w})$ converges to $\nabla^{co\mathcal{E}uc^3}_{\sigma v}\sigma w$ by Proposition 5.14;
- $g_t^*\nabla^{\mathrm{I}_t}_v(\hat{w})$ converges to a tangential component to $\sigma(S)$, since $\nabla^{\mathrm{I}_t}_v(\hat{w})$ is tangent to $\sigma_t(S)$;
- $tg_t^*(N_t)$ converges to $\mathrm{T} = (0, 0, 0, 1)$ since $g_t^*(\mathrm{N}_t) = (\mathrm{N}_t^1, \mathrm{N}_t^2, \mathrm{N}_t^3, \mathrm{N}_t^4/t)$ and $\mathrm{N}_0 = (0, 0, 0, 1)$.

This shows at once that $\mathrm{II}_t(v,w)/t$ converges to the second fundamental form of σ in $co\mathcal{E}uc^3$, and $g_t^*\nabla^{\mathrm{I}_t}_v(\hat{w})$ converges to $\nabla^{\mathrm{I}}_v(\hat{w})$.

Since $\mathrm{II}_t(v,w) = \mathrm{I}_t(B_t(v), w)$ and $\mathrm{II}(v,w) = \mathrm{I}(B(v), w)$, the third point follows from the first two statements. The last point is a consequence of the third point and the fact that $K_t^{\mathrm{ext}} = \det B_t$ and $K^{\mathrm{ext}} = \det B$.

Clearly the proof is completely analogous for the convergence from $\mathrm{d}\mathbb{S}^3$ to ${}^*\mathbb{E}^3$, or from the convergence from $\mathbb{H}^3$ or $\mathbb{A}\mathrm{d}\mathbb{S}^3$ to ${}^*\mathbb{M}\mathbb{i}\mathbb{n}^3$. □

6.5 Projective nature of infinitesimal rigidity

Let S be a smooth space-like surface in $\mathbb{E}^3$, $\mathbb{M}\mathrm{in}^3$, $\mathbb{E}\mathrm{ll}^3$, $\mathbb{H}^3$, $\mathrm{d}\mathbb{S}^3$, or $\mathbb{A}\mathrm{d}\mathbb{S}^3$. A vector field Z on S is called an *infinitesimal isometric deformation* if it preserves the induced metric on S at the first order: $\mathcal{L}_Z g(X,Y) = 0$ for all $X, Y \in TS$, which is easily shown to be equivalent to $g(\nabla_X Z, X) = 0$ for all $X \in TS$, if g is the metric of the ambient space and ∇ its Levi–Civita connection. An infinitesimal isometric deformation is said to be *trivial* if it is the restriction to S of a Killing field of the ambient space.

Now let us suppose that S embeds in an affine chart, and that this affine representation is also an affine chart for another model space, in which the image of S is space-like. By a straightforward adaptation of the proof of Lemma 5.18, it is easy to see that the infinitesimal Pogorelov map sends infinitesimal isometric deformations of S for the first ambient metric to infinitesimal isometric deformations of S for the second ambient metric. Moreover, Lemma 5.18 also says that trivial deformations are sent to trivial deformations. A surface S is *infinitesimally rigid* if all its infinitesimal isometric deformations are trivial.

It follows from the above discussion that the infinitesimal rigidity of a surface in a three dimensional affine space is independent of the choice of the projective distance on the ambient space.

Comments and references

- A classical problem in Riemannian geometry is the existence of zero mean curvature surfaces with prescribed boundary at infinity. For instance, in hyperbolic space the so-called *asymptotic Plateau problem* was proved by Anderson in [5]. In anti-de Sitter space, the analogous problem was tackled by Bonsante and Schlenker in [16]. With the tools introduced in this paper, the analogous problem in $^*\mathbb{M}\mathrm{in}^3$ turns out to be very simple. Indeed, taking the trace in the expression $B = \mathrm{Hess}^{\mathbb{H}^3} u - u\,\mathrm{Id}$ for the shape operator of a space-like surface, one obtains that a surface has zero mean curvature if and only if $\Delta u - 2u = 0$ where Δ denotes the hyperbolic Laplacian. Hence a proof follows from the existence and uniqueness of solutions to this linear PDE.

- Volkov [68] showed the following: a vector field Z on S is an infinitesimal isometric deformation if and only if for each $x \in S$ there exists a Killing field K^x such that $Z_x = (K^x)_x$ and $(\nabla_X Z)_x = (\nabla_X K^x)_x$ for all $X \in T_x S$.

- To simplify the presentation, we have considered different metrics on the same (affine) space. But we could have considered maps between different spaces, sending geodesics onto geodesics. In particular, it would follow from the result presented here that infinitesimal rigidity in the Euclidean space is invariant under projective transformations. This fact is known since the end of the 19th century and called the Darboux–Sauer theorem, see e.g. [31, Chapitre IV]. This is also true for more general surfaces than smooth ones. For example consider polyhedral surfaces made of pieces of space-like planes. An infinitesimal isometric deformation is then defined as the data of a Killing field on each face, such that they coincide on common edges. Then Lemma 5.18 implies that the infinitesimal Pogorelov map sends deformations to deformations and trivial deformations to trivial deformations. This invariance of infinitesimal rigidity is always true if one considers the more general case of frameworks, that includes polyhedral surfaces [39, 38].

- The *Pogorelov map* is a map taking two convex surfaces S_1, S_2 in $\mathbb{E}\mathbb{l}\mathbb{l}^3$ and giving two convex surfaces $\overline{S}_1, \overline{S}_2$ of $\mathbb{E}^3$. If S_1 is isometric to S_2 (for the distance induced by the ambient metric) then $\overline{S}_1$ and $\overline{S}_2$ are also isometric. Moreover, if the isometry between S_1 and S_2 is the restriction to S_1 of an isometry of the ambient space, then the isometry between $\overline{S}_1$ and $\overline{S}_2$ is also the restriction of an isometry of the ambient space. The difficult part is to prove that $\overline{S}_1$ and $\overline{S}_2$ are convex [51]. The differential of this map at $S = S_1 = S_2$ gives the infinitesimal Pogorelov map. The map was extended to $\mathbb{H}^3$ [51] and to $\mathrm{d}\mathbb{S}^3$ [47] instead of $\mathbb{E}\mathbb{l}\mathbb{l}^3$, see also [56, 57].

References

[1] N. A'Campo and A. Papadopoulos, "Notes on non-Euclidean geometry," in *Strasbourg master class on geometry* (A. Papadopoulos, ed.), IRMA Lectures in Mathematics and Theoretical Physics 18, European Mathematical Society (EMS), Zürich, 2012, 1–182.
MR 2931886 Zbl 1253.51001 ☞ 321

[2] N. A'Campo and A. Papadopoulos, "On Klein's 'So-called non-Euclidean geometry,'," in *Sophus Lie and Felix Klein*: *The Erlangen program and its impact in mathematics and in physics* (L. Ji and A. Papadopoulos, eds.), IRMA Lectures in Mathematics and Theoretical Physics 23, European Mathematical Society (EMS), Zürich, 2015. 91–136.
MR 3362214 ☞ 321, 339

[3] N. A'Campo and A. Papadopoulos, "Transitional geometry," in *Sophus Lie and Felix Klein*: *The Erlangen program and its impact in mathematics and in physics* (L. Ji and A. Papadopoulos, eds.), IRMA Lectures in Mathematics and Theoretical Physics 23, European Mathematical Society (EMS), Zürich, 2015. 217–235.
MR 3362217 ☞ 372

[4] I. Agol. "The virtual Haken conjecture," *Doc. Math.*, 18:1045–1087, 2013. with an appendix by I. Agol, D. Groves, and J. Manning, *Doc. Math.* 18 (2013), 1045–1087.
MR 3104553 Zbl 1286.57019 ☞ 322

[5] M. T. Anderson, "Complete minimal hypersurfaces in hyperbolic n-manifolds," *Comment. Math. Helv.* **58** (1983), no. 2, 264–290.
MR 0705537 Zbl 0549.53058 ☞ 400

[6] L. Andersson, T. Barbot, R. Benedetti, F. Bonsante, W. M. Goldman, F. Labourie, K. P. Scannell, and J.-M. Schlenker, "Notes on: 'Lorentz spacetimes of constant curvature' by G. Mess," *Geom. Dedicata* 126 (2007), 47–70.
MR 2328922 Zbl 1126.53042 ☞ 322

[7] C. Bär, P. Gauduchon, and A. Moroianu, "Generalized cylinders in semi-Riemannian and Spin geometry," *Math. Z.* 249 (2005), no. 3, 545–580.
MR 2121740 Zbl 1068.53030 ☞ 390

[8] T. Barbot, F. Béguin, and A. Zeghib, "Constant mean curvature foliations of globally hyperbolic spacetimes locally modelled on AdS_3," *Geom. Dedicata* 126 (2007), 71–129.
MR 2328923 Zbl 1255.83118 ☞ 322

[9] T. Barbot, F. Béguin, and A. Zeghib, "Prescribing Gauss curvature of surfaces in 3-dimensional spacetimes: application to the Minkowski problem in the Minkowski space," *Ann. Inst. Fourier* (*Grenoble*) 61 (2011), no. 2, 511–591.
MR 2895066 Zbl 1234.53019 ☞ 322

[10] J. K. Beem, P. E. Ehrlich, and K. L. Easley, *Global Lorentzian geometry,* Second edition. Monographs and Textbooks in Pure and Applied Mathematics, 202. Marcel Dekker, New York, 1996.
MR 1384756 Zbl 0846.53001 ☞ 322

[11] R. Benedetti and C. Petronio. *Lectures on hyperbolic geometry,* Universitext. Springer-Verlag, Berlin etc., 1992.
MR 1219310 Zbl 0768.51018 ☞ 322

[12] I. Bivens, J.-P. Bourguignon, A. Derdziński, D. Ferus, O. Kowalski, T. Klotz Milnor, V. Oliker, U. Simon, W. Strübing, and K. Voss, “Discussion on Codazzi-tensors,” in *Global differential geometry and global analysis* (D. Ferus, W. Kühnel, U. Simon, and B. Wegner, eds.), Proceedings of the Colloquium held at the Technical University of Berlin, Berlin, November 21–24, 1979, Lecture Notes in Mathematics 838, Springer, Berlin etc., 1981, 243–299.
MR 0636289 ☞ 394

[13] F. Bonahon and J.-P. Otal, “Laminations measurées de plissage des variétés hyperboliques de dimension 3,” *Ann. of Math.* (2), 160 (2004), no. 3, 1013–1055.
☞ 322

[14] F. Bonsante and F. Fillastre, “The equivariant Minkowski problem in Minkowski space,” *Ann. Inst. Fourier* (*Grenoble*) 67 (2017), no. 3, 1035–1113.
MR 3668768 Zbl 06821969 356, 358

[15] F. Bonsante, G. Mondello, and J.-M. Schlenker, “A cyclic extension of the earthquake flow I,” *Geom. Topol.* 17 (2013), no. 1, 157–234.
MR 3035326 Zbl 1359.37079 ☞ 322

[16] F. Bonsante and J.-M. Schlenker, “Maximal surfaces and the universal Teichmüller space,” *Invent. Math.* 182 (2010), no. 2, 279–333.
MR 2729269 Zbl 1222.53063 ☞ 322, 400

[17] F. Bonsante and J.-M. Schlenker, “Fixed points of compositions of earthquakes,” *Duke Math. J.* 161 (2012), no. 6, 1011–1054.
MR 2913100 Zbl 1244.32007 ☞ 322

[18] F. Bonsante and A. Seppi, “Equivariant maps into anti-de Sitter space and the symplectic geometry of $\mathbb{H}^2 \times \mathbb{H}^2$,” to appear in *Trans. Amer. Math. Soc.*
☞ 322

[19] F. Bonsante and A. Seppi, “On Codazzi tensors on a hyperbolic surface and flat Lorentzian geometry,” *Int. Math. Res. Not. IMRN* 2016, no. 2, 343–417.
MR 3493421 Zbl 1336.53034 ☞ 322

[20] F. Bonsante and A. Seppi, Spacelike convex surfaces with prescribed curvature in $(2+1)$-Minkowski space. *Adv. Math.* 304 (2017), 434–493.
MR 3558215 Zbl 1351.53068 ☞ 322

[21] F. Bonsante and A. Seppi, Area-preserving diffeomorphism of the hyperbolic plane and K-surfaces in anti-de Sitter space. *J. Topol.* 11 (2018), no. 2, 420–468.
MR 3789829 ☞ 322

[22] F. Bonsante, A. Seppi, and A. Tamburelli, “On the volume of anti-de Sitter maximal globally hyperbolic three-manifolds,” *Geom. Funct. Anal.* 27 (2017), no. 5, 1106–1160.
MR 3714718 Zbl 1390.53075 ☞ 322

[23] Y. Cho, “Trigonometry in extended hyperbolic space and extended de Sitter space,” *Bull. Korean Math. Soc.* 46 (2009), no. 6, 1099–1133.
MR 2583477 Zbl 1194.51022 ☞ 341

[24] D. Cooper, J. Danciger, and A. Wienhard, “Limits of geometries,” *Trans. Amer. Math. Soc.* 370 (2018), no. 9, 6585–6627.
MR 3814342 Zbl 06891720 ☞ 358, 372

[25] D. Cooper, C. D. Hodgson, and S. P. Kerckhoff, *Three-dimensional orbifolds and cone-manifolds,* with a postface by S. Kojima, MSJ Memoirs 5, Mathematical Society of Japan, Tokyo, 2000.
MR 1778789 Zbl 0955.57014 ☞ 322

[26] J. Danciger, *Geometric transitions: From hyperbolic to AdS geometry,* Ph.D. thesis, Stanford University, Stanford CA, 2011.
☞ 322

[27] J. Danciger, “A geometric transition from hyperbolic to anti-de Sitter geometry,” *Geom. Topol.* 17 (2013), no. 5, 3077–3134.
MR 3190306 Zbl 1287.57020 ☞ 322

[28] J. Danciger, F. Guéritaud, and F. Kassel, “Geometry and topology of complete Lorentz spacetimes of constant curvature,” *Ann. Sci. Éc. Norm. Supér.* (4) 49 (2016), no. 1, 1–56.
MR 3465975 Zbl 1344.53049 ☞ 322

[29] J. Danciger, F. Guéritaud, and F. Kassel, “Margulis spacetimes via the arc complex,” *Invent. Math.* 204 (2016), no. 1, 133–193.
MR 3480555 Zbl 1344.30035 ☞ 322

[30] J. Danciger, S. Maloni, and J.-M. Schlenker, “Polyhedra inscribed in a quadric,” Preprint, 2014.
arXiv:1410.3774 [math.DG]
☞ 359, 387

[31] G. Darboux, *Leçons sur la théorie générale des surfaces et les applications géométriques du calcul infinitésimal,* IV$^{\text{e}}$ Partie, Déformation infiniment petite et représentation sphérique. Chelsea Publishing Co., Bronx, N.Y., 1972. Déformation infiniment petite et representation sphérique, Reprint of the first edition, Gauthier-Villars et Fils, Paris, 1896.
JFM 25.1159.02 (first edition) Zbl 0257.53001 (reprint) ☞ 401

[32] B. Deroin and N. Tholozan, "Dominating surface group representations by Fuchsian ones," *Int. Math. Res. Not. IMRN* 2016, no. 13, 4145–4166.
MR 3544632 ☞ 322

[33] D. Gale, V. Klee, and R. T. Rockafellar, "Convex functions on convex polytopes," *Proc. Amer. Math. Soc.* 19 (1968), 867–873.
MR 0230219 Zbl 0246.26009 ☞ 359

[34] R. F. Galeeva, V. V. Pelipenko, and D. D. Sokolov, "Infinitesimal bendings of surfaces in pseudo-Euclidean space," *Uspekhi Mat. Nauk* 37 (1982), no. 5(227), 175–176, in Russian. English translation, *Russian Math. Surveys* 37 (1982), no. 5, 172–173.
MR 0676623 Zbl 0508.53015 ☞ 387

[35] S. Gallot, D. Hulin, and J. Lafontaine, *Riemannian geometry,* third edition, Universitext, Springer-Verlag, Berlin, 2004.
MR 2088027 Zbl 1068.53001 ☞ 394

[36] S. W. Hawking and G. F. R. Ellis, *The large scale structure of space-time,* Cambridge Monographs on Mathematical Physics 1, Cambridge University Press, London and New York, 1973.
MR 0424186 Zbl 0265.53054 ☞ 322

[37] C. D. Hodgson and S. P. Kerckhoff, "Rigidity of hyperbolic cone-manifolds and hyperbolic Dehn surgery," *J. Differential Geom.* 48 (1998), no. 1, 1–59.
MR 1622600 Zbl 0919.57009 ☞ 322

[38] I. Izmestiev, "Statics and kinematics of frameworks in Euclidean and non-Euclidean geometry," in *Eighteen essays in non-Euclidean geometry* (V. Alberge and A. Papadopoulos, eds.), European Mathematical Society (EMS), Zürich, 2019, Chapter 12, 191–233.
☞ 401

[39] I. Izmestiev, "Projective background of the infinitesimal rigidity of frameworks," *Geom. Dedicata* 140 (2009), 183–203.
MR 2504741 Zbl 1169.52010 ☞ 401

[40] M. S. Knebelman, "On groups of motion in related spaces," *Amer. J. Math.* 52 (1930), no. 2, 280–282.
MR 1506754 JFM 56.1197.02 ☞ 387

[41] S. Kobayashi, *Transformation groups in differential geometry,* reprint of the 1972 edition, Classics in Mathematics, Springer-Verlag, Berlin, 1995.
MR 1336823 Zbl 0829.53023 ☞ 387

[42] K. Kozai, "Hyperbolic structures from Sol on pseudo-Anosov mapping tori," *Geom. Topol.* 20 (2016), no. 1, 437–468.
MR 3470718 Zbl 1339.57025 ☞ 322, 372

[43] K. Krasnov and J.-M. Schlenker, "Minimal surfaces and particles in 3-manifolds," *Geom. Dedicata* 126 (2007), 187–254.
MR 2328927 Zbl 1126.53037 ☞ 390

[44] F. Labourie, "Surfaces convexes dans l'espace hyperbolique et $\boldsymbol{C}\mathrm{P}^1$-structures," *J. London Math. Soc.* (2) 45 (1992), no. 3, 549–565.
MR 1180262 Zbl 0767.53011 ☞ 390

[45] F. Labourie and J.-M. Schlenker, "Surfaces convexes fuchsiennes dans les espaces lorentziens à courbure constante," *Math. Ann.* 316 (2000), no. 3, 465–483.
MR 1752780 Zbl 0968.53047 ☞ 387

[46] G. Mess, "Lorentz spacetimes of constant curvature," *Geom. Dedicata* 126 (2007), 3–45.
MR 2328921 Zbl 1206.83117 ☞ 322

[47] A. D. Milka, Unique determination of convex surfaces in a pseudo-Riemannian spherical space. I. *Ukrain. Geom. Sb.* 29 (1986), 113–118, iii, in Russian. English translation, *J. Soviet Math.* 51 (1990), no. 2, 2273–2276. MR 0866306 Zbl 0607.53037 ☞ 401

[48] K. Nomizu and T. Sasaki, *Affine differential geometry,* Geometry of affine immersions, Cambridge Tracts in Mathematics 111. Cambridge University Press, Cambridge, 1994.
MR 1311248 Zbl 0834.53002 ☞ 386

[49] B. O'Neill, *Semi-Riemannian geometry,* With applications to relativity, Pure and Applied Mathematics 103, Academic Press, New York, 1983.
MR 0719023 Zbl 0531.53051 ☞ 322, 325

[50] A. Papadopoulos, *Metric spaces, convexity and non-positive curvature,* second edition, IRMA Lectures in Mathematics and Theoretical Physics 6, European Mathematical Society (EMS), Zürich, 2014.
MR 3156529 Zbl 1296.53007 ☞ 339

[51] A. V. Pogorelov, *Внешняя геометрия выпуклых поверхностей,* Izdat. “Nauka,” Moscow, 1969, in Russian. English translation, *Extrinsic geometry of convex surfaces,* Translated by Israel Program for Scientific Translations, Translations of Mathematical Monographs 35, American Mathematical Society, Providence, R.I., 1973.
MR 0244909 MR0346714 (transl.) Zbl 0311.53067 ☞ 401

[52] H. Poincaré, “Sur les hypothèses fondamentales de la géométrie,” *Bull. Soc. Math. France* 15 (1887), 203–216.
MR 1504026 JFM 19.0512.01 ☞ 339

[53] J. Richter-Gebert, *Perspectives on projective geometry,* A guided tour through real and complex geometry. Springer, Heidelberg, 2011.
MR 2791970 Zbl 1214.51001 ☞ 329, 358

[54] R. T. Rockafellar, *Convex analysis,* reprint of the 1970 original, Princeton Landmarks in Mathematics, Princeton Paperbacks, Princeton University Press, Princeton, N.J., 1997.
MR 1451876 Zbl 0932.90001 ☞ 353

[55] B. A. Rosenfeld, *A history of non-Euclidean geometry,* Evolution of the concept of a geometric space, translated from the Russian by A. Shenitzer, Studies in the History of Mathematics and Physical Sciences 12, Springer-Verlag, New York, 1988.
MR 0959136 Zbl 0657.01002 ☞ 321, 322

[56] J.-M. Schlenker, “Métriques sur les polyèdres hyperboliques convexes,” *J. Differential Geom.* 48 (1998), no. 2, 323–405.
MR 1630178 Zbl 0912.52008 ☞ 401

[57] J.-M. Schlenker, “Dihedral angles of convex polyhedra,” *Discrete Comput. Geom.* 23 (2000), no. 3, 409–417.
MR 1744513 Zbl 0951.52006 ☞ 401

[58] J.-M. Schlenker, “Hyperbolic manifolds with convex boundary,” *Invent. Math.* 163 (2006), no. 1, 109–169.
MR 2208419 Zbl 2208419 ☞ 387

[59] R. Schneider, *Convex bodies*: *the Brunn–Minkowski theory,* second expanded edition, Encyclopedia of Mathematics and its Applications 151, Cambridge University Press, Cambridge, 2014.
MR 3155183 Zbl 1287.52001 ☞ 358

[60] A. Seppi, *Surfaces in constant curvature three-manifolds and the infinitesimal Teichmüller theory,* Ph.D. thesis, Università degli studi di Pavia, Pavia, 2015.
☞ 322

[61] A. Seppi, "Maximal surfaces in anti-de Sitter space, width of convex hulls and quasiconformal extensions of quasisymmetric homeomorphisms," to appear in *J. Eur. Math. Soc.* (*JEMS*).
☞ 322

[62] A. Seppi, "Minimal discs in hyperbolic space bounded by a quasicircle at infinity," *Comment. Math. Helv.* 91 (2016), no. 4, 807–839.
MR 3566524 Zbl 1356.53063 ☞ 322

[63] M. Spivak, *A comprehensive introduction to differential geometry,* Vol. III, second edition. Publish or Perish, Wilmington, DE, 1979.
MR 0532832 Zbl 0439.53003 ☞ 390

[64] H. Struve and R. Struve, "Projective spaces with Cayley–Klein metrics," *J. Geom.* 81 (2004), no. 1-2, 155–167.
MR 2134074 Zbl 1078.51001 ☞ 322, 358

[65] W. P. Thurston, *The geometry and topology of three-manifolds,* mimeographed notes, Princeton University, Princeton, N.J., 1979.
`http://library.msri.org/books/gt3m/`
☞ 321, 340, 341

[66] W. P. Thurston, *Three-dimensional geometry and topology,* Vol. 1, edited by S. Levy, Princeton Mathematical Series 35, Princeton University Press, Princeton, N.J., 1997.
MR 1435975 Zbl 0873.57001 ☞ 321, 340

[67] W. P. Thurston, "Earthquakes in 2-dimensional hyperbolic geometry," in *Fundamentals of hyperbolic geometry: selected expositions* (R. D. Canary, D. Epstein and A. Marden, eds.), London Mathematical Society Lecture Note Series 328, Cambridge University Press, Cambridge, 2006, 267–289.
MR 2235712 Zbl 0628.57009 ☞ 322

[68] J. A. Volkov, “Generalization of theorems of Darboux-Sauer and Pogorelov,” In *Вопросы глобальной геометрии* (*Problems in global geometry,* Ju. D. Burago, ed.) Zap. Naučn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI) 45 (1974). Izdat. “Nauka” Leningrad. Otdel., Leningrad, 1974, 63–67, 118, in Russian. English translation, *J. Sov. Math.* 8 (1978), 444–448
MR 0388295 Zbl 0399.53006 ☞ 387, 400

[69] I. M. Yaglom, *A simple non-Euclidean geometry and its physical basis,* An elementary account of Galilean geometry and the Galilean principle of relativity, Heidelberg Science Library, translated from the Russian by A. Shenitzer, with the editorial assistance of B. Gordon, Springer-Verlag New York, New York, 1979.
MR 0520230 Zbl 0393.51013 ☞ 371

PART III
Other geometries

Chapter 17

Hermitian trigonometry

Boumediene Et-Taoui

Contents

1. Introduction 413
2. Congruence theorem for triangles in $\mathbb{C}P^2$ 414

References 423

1 Introduction

The aim of this chapter is to present the trigonometry of complex projective spaces. Section 2 is devoted to a congruence theorem for triangles in $\mathbb{C}P^2$ obtained by Brehm. We prove it here by another method which we presented in [3] and in [7]. It is shown that a triangle $([x], [y], [z])$ in $\mathbb{C}P^2$ with $\|x\| = \|y\| = \|z\| = 1$ is uniquely determined up to isometry by its side lengths $a = \arccos(|\langle y, z\rangle|)$, $b = \arccos(|\langle x, z\rangle|)$ and $c = \arccos(|\langle x, y\rangle|)$ and its *shape invariant*, the real part of the complex number

$$\langle x, y\rangle \langle y, z\rangle \langle z, x\rangle = \cos(a) \cos(b) \cos(c) e^{i\omega}.$$

Section 3 consists of Hermitian trigonometry. as it is developed in [4]. Precisely, we give the two laws of sines and the law of cosines of a triangle in $\mathbb{C}P^2$. In Section 4, we define the polar triangle associated to a given triangle and derive from it the second law of sines which does not appear in [2]. On the other hand, if we consider a triangle $([x], [y], [z])$ and its polar triangle $([X], [Y], [Z])$, and if we set

$$\langle X, Y\rangle \langle Y, Z\rangle \langle Z, X\rangle = \cos(p) \cos(q) \cos(r) e^{i\Omega},$$

we show that ω and Ω are linked. Indeed, consider the complex number

$$\langle u_2, u_1\rangle \langle v_1, v_2\rangle \langle w_1, w_2\rangle = |\langle u_2, u_1\rangle \langle v_1, v_2\rangle \langle w_1, w_2\rangle| e^{i\alpha},$$

where $u_1, u_2; v_1, v_2; w_1, w_2$ are pairs of unit tangent vectors at the vertex $[x]$, $[y]$, $[z]$ respectively. It is displayed that

$$\Omega = -\alpha - 2\omega + (2k+1)\pi, \quad k \in \mathbb{Z}.$$

This result is new in comparison with the results in the papers quoted. In addition, we relate Brehm's invariants to those of Hsiang [10] and those of Shirokov and Rozenfeld [14]. We close this paper by some remarks on trigonometry in general symmetric spaces.

2 Congruence theorem for triangles in $\mathbb{C}P^2$

In classical geometry, Euclidean, spherical, and hyperbolic spaces are commonly characterized by the same set of congruence axioms for geodesic intervals and triangles. In the language of Riemannian geometry, these axioms are stated as follows ([10]):

1. two geodesic intervals AB and $A'B'$ are congruent if they have the same length;
2. two geodesic triangles are congruent if they have two equal angles contained by a pair of equal sides.

Let ABC be a geodesic triangle in a symmetric space M of rank 1 and $\Delta(M)$ the moduli space of congruence classes of triangles in M. In the well-known classical spaces, a triangle has six basic invariants (side lengths and angles) which map $\Delta(M)$ into $\mathbb{R}^6$. The image set is of dimension 3 since the 6 invariants satisfy three basic functional relations, the laws of sines (which involve 2 equalities) and cosines (one equality). Riemannian symmetric spaces of rank 1 are precisely the two-point homogeneous spaces [18].

In 1994, in his Ph.D. thesis [7], the author gave new congruence criteria for finite subsets in both complex projective and complex hyperbolic spaces. The case of triangles is then easily derived. Using this point of view, he derived an easy congruence theorem for triangles. This leads directly to the laws of trigonometry in $\mathbb{C}P^2$ obtained by Brehm, from which one can easily derive those of the sphere.

We consider points in $\mathbb{C}P^n$, $n \geq 2$ as one-dimensional subspaces of the vector space $\mathbb{C}^{n+1}$. For $x \in \mathbb{C}^{n+1}$, $x \neq 0$, let $[x] := \mathbb{C} \cdot x$ denote the corresponding point in $\mathbb{C}P^n$. Let $\langle x, y\rangle = \sum_{i=1}^{n+1} \overline{x_i} y_i$ be the usual Hermitian product and let d be the Fubini–Study metric on $\mathbb{C}P^n$ defined by

$$d([x],[y]) = \arccos\left(\frac{|\langle x, y\rangle|}{\|x\|\,\|y\|}\right).$$

It is well known that $\Phi\colon \mathbb{C}P^n \to \mathbb{C}P^n$ is an isometry of $\mathbb{C}P^n$ if and only if there exists a unitary matrix U such that either $\Phi([x]) = [Ux]$ for all $x \in \mathbb{C}^{n+1}$ or $\Phi([x]) = [\overline{Ux}]$ for all $x \in \mathbb{C}^{n+1}$.

Let $[x], [y], [z] \in \mathbb{C}P^n$ with $\|x\| = \|y\| = \|z\| = 1$.

Definition 2.1. A *triangle* is a triple $([x], [y], [z])$ of three different points in $\mathbb{C}P^n$ such that the distance between any two of them is strictly smaller than $\frac{\pi}{2}$.

We imagine the vertices $[x]$, $[y]$, $[z]$ of a triangle being connected by the unique shortest paths between them, called sides. We set

$$a = \arccos(|\langle y, z\rangle|), \quad b = \arccos(|\langle x, z\rangle|), \quad c = \arccos(|\langle x, y\rangle|).$$

Consider the Hermitian matrix

$$H = \begin{pmatrix} 1 & \langle x, y\rangle & \langle x, z\rangle \\ \langle y, x\rangle & 1 & \langle y, z\rangle \\ \langle z, x\rangle & \langle z, y\rangle & 1 \end{pmatrix},$$

which is positive semidefinite with 1's on the diagonal. This matrix can be brought, by multiplication of columns by unimodular complex numbers and the corresponding rows by the conjugate of those complex numbers into the following form

$$\begin{pmatrix} 1 & \cos(c) & \cos(b) \\ \cos(c) & 1 & \cos(a)e^{i\omega} \\ \cos(b) & \cos(a)e^{-i\omega} & 1 \end{pmatrix}. \tag{1}$$

The complex number $e^{i\omega}$ is called the *direct shape invariant* in [3]. It was considered by Blaschke and Terheggen. Terheggen interpreted it in spherical geometry [15]. The real number $\cos\omega$ is called by Brehm the *shape invariant.* It was shown by Brehm [4] that a triangle in $\mathbb{C}P^n$ is determined, up to isometry, by its side lengths and its shape invariant. Brehm used the invariant $\sigma = \cos(a)\cos(b)\cos(c)\cos(\omega)$. The author showed in his Ph.D. thesis that two triangles in $\mathbb{C}P^n$ are directly congruent (if we do not allow conjugation) if and only if they have equivalent associated matrices. We already saw that the form (1) is determined up to equivalence by the real numbers a, b, c, and $\cos(\omega)$ (conjugation is allowed). Now let H be any Hermitian positive semidefinite 3-matrix with 1's on the diagonal. Clearly it can be decomposed as the product T^*T where T is an upper triangular matrix $(n+1, 3)$, $n \geq 2$ with diagonal positive real numbers and T^* is the transpose of its conjugate. The columns of T are the representing vectors of the desired triangle.

More general results on finite subsets in complex projective spaces can be found in [3].

A straightforward calculation gives the following vectors which are the columns of the upper triangular matrix T,

$$\begin{aligned} x &= (1,0,0,\dots,0),\\ y &= (\cos(c),\sin(c),0,\dots,0),\\ z &= (\cos(b), z_2+z_2' i, z_3, 0,\dots,0), \end{aligned}$$

where

$$z_2 = \frac{\cos(a)\cos(\omega)-\cos(c)\cos(b)}{\sin(c)}, \tag{2}$$

$$z_2' = \frac{\cos(a)\sin(\omega)}{\sin(c)}, \tag{3}$$

$$z_3 = \frac{(1-\cos^2(a)-\cos^2(b)-\cos^2(c)+2\cos(a)\cos(b)\cos(c)\cos(\omega))^{1/2}}{\sin(c)}. \tag{4}$$

The triple $([x],[y],[z])$ is called in [4] the normal form of the congruence class of the triangle. We have then proved the following theorem due to Brehm. The proof presented here is slightly different from Brehm's one and was obtained in the author's Ph.D. thesis [7] and published in [3].

Theorem 2.2. 1. *A triangle in* $\mathbb{C}P^n$ *is determined uniquely up to isometry by its side lengths and its shape invariant.*

2. *Let* $0 < a, b, c < \frac{\pi}{2}$ *and* $\sigma \in \mathbb{R}$. *Then there exists a triangle in* $\mathbb{C}P^n$ *with side lengths* a, b, c *and shape invariant* σ *if and only if*

$$\frac{1}{2}(\cos^2(a)+\cos^2(b)+\cos^2(c)-1) \le \sigma \le |\sigma| \le \cos(a)\cos(b)\cos(c).$$

3 Complex projective trigonometry

Let $x \in \mathbb{C}^{n+1}$, $\|x\| = 1$. The tangent space of $\mathbb{C}P^n$ at $[x]$ can be canonically identified with $(\mathbb{C}\cdot x)^{\perp}$ in $\mathbb{C}^{n+1}$. The real part $\Re(\langle,\rangle)$ of the Hermitian product is the Riemannian metric of $\mathbb{C}P^n$. We represent as in [4] the tangent unit vector $u(x,[y])$ at $[x]$ of the shortest geodesic from $[x]$ to $[y]$ as

$$u(x,[y]) := \frac{y\,\langle y,x\rangle - x\,|\langle y,x\rangle|^2}{\|y\,\langle y,x\rangle - x\,|\langle y,x\rangle|^2\,\|}.$$

Consider the normal form given above and let us set $u_1 = u(x, [y])$ and $u_2 = u(x, [z])$. By a straightforward calculation one has

$$u_1 = (0, 1, 0, \dots, 0), \quad u_2 = (0, z_2 + z_2' i, z_3, 0, \dots, 0) \sin^{-1}(b).$$

The angular invariants of [4] are defined by

$$\Lambda(A) := \arccos(\Re(\langle u_1, u_2 \rangle)),$$

$$\Phi(A) := \arccos(|\langle u_1, u_2 \rangle|),$$

$$\Psi(A) := \arcsin(|\Im(\langle u_1, u_2 \rangle)|),$$

where A, B, C correspond to the corners $[x]$, $[y]$, $[z]$ respectively. Geometrically, $\Lambda(A)$ is the angle between the geodesics, $\Phi(A)$ is the angle between the projective lines generated by u_1 and u_2 and $\Psi(A)$ is the smallest angle between the geodesic from $[x]$ to $[z]$ and a totally geodesic $\mathbb{R}P^2$ containing $[x]$ and $[y]$. However, the three angles are dependent, they satisfy $\sin^2 \Lambda(A) = \sin^2(\Phi(A)) + \sin^2(\Psi(A))$. Clearly the pair of angular invariants of one corner together with the lengths of its adjacent sides constitute a minimal set of congruence conditions. Thus, we have $\dim \Delta(\mathbb{C}P^n) = 4$. Therefore there exist five basic functional relations among the nine invariants of a triangle.

A short computation yields

$$\cos(\Lambda(A)) = \frac{z_2}{\sin(b)}, \tag{5}$$

$$\sin(\Psi(A)) = \frac{z_2'}{\sin(b)}, \tag{6}$$

$$\sin \Phi(A) = \frac{z_3}{\sin(b)}, \tag{7}$$

Theorem 3.1. *For any triangle ABC in $\mathbb{C}P^n$,*

$$\frac{\sin(\Phi(A))}{\sin(a)} = \frac{\sin(\Phi(B))}{\sin(b)} = \frac{\sin(\Phi(C))}{\sin(c)}, \tag{8}$$

$$\frac{\sin(\Psi(A))}{\sin(2a)} = \frac{\sin(\Psi(B))}{\sin(2b)} = \frac{\sin(\Psi(C))}{\sin(2c)}, \tag{9}$$

$$\begin{aligned} &\cos^2(a) - (\cos(b)\cos(c) + \sin(b)\sin(c)\cos(\Lambda(A)))^2 \\ &\quad = \sin^2(\Psi(A))\sin^2(b)\sin^2(c). \end{aligned} \tag{10}$$

Proof. Equation (8) is obtained using (4) and (7). Equation (9) is obtained from (3) and (6). Now for the law of cosines we have from (2) and (5),

$$\begin{aligned}&\cos(a)\cos(b)\cos(c)\cos(\omega)\\&\quad=\cos(b)\cos(c)\sin(b)\sin(c)\cos(\Lambda(A))+\cos^2(b)\cos^2(c),\end{aligned}\tag{11}$$

and from (4) and (7),

$$\begin{aligned}&2\cos(a)\cos(b)\cos(c)\cos(\omega)\\&\quad=\sin^2(\Phi(A))\sin^2(c)\sin^2(b)-1+\cos^2(a)+\cos^2(b)+\cos^2(c).\end{aligned}\tag{12}$$

Combining (11), (12), and the relation between Λ, Φ and Ψ we arrive easily to the law of cosines (10). ☐

The law of cosines (10) does not appear in this form in Hsiang's and Brehm's articles. From (10) we find again the well-known law of cosines in elliptic and spherical spaces.

Brehm's law of cosines reads as follows:

$$\begin{aligned}&\cos(2b)\cos(2c)-\cos(2a)+\cos(\Lambda(A))\sin(2b)\sin(2c)\\&\quad=2\sin^2(\Phi(A))\sin^2(b)\sin^2(c).\end{aligned}\tag{13}$$

Remark 3.2. 1. If $\Phi(A)=0$ then (13) is exactly the classical cosine law for spherical triangles, noting that $\mathbb{C}P^1$ is isometric to the sphere of radius $\frac{1}{2}$.

2. If $\Psi(A)=0$ then (10) is exactly the classical cosine law for elliptic and thus spherical triangles.

3. The congruence theorem for geodesic triangles was stated for the first time by Blaschke and Terheggen [2]. They showed that the congruence class of a geodesic triangle in a complex projective plane is determined by its side lengths and a fourth invariant $\cos\phi$, satisfiying

$$\cos(\Phi(A))\sin(\phi)=\sin(\Psi(A)).$$

Later, Terheggen interpreted in [15] the invariant ϕ as an angle in spherical geometry. In addition, he showed that all spherical trigonometry in the sense of Study [16] is contained in the geometry of the complex projective line. Although Blaschke and Terheggen were unaware that we need five relations, they obtained three. By means of the notion of polar triangle in a complex projective plane, they obtained a sine and a cosine law. However, it is seen below that the second sine law can be easily derived from their investigation. Their sine law leads easily to the well-know spherical sine law obtained by Coolidge [5]. Another approach was put forward by Shirokov, in a paper published posthumously by Rozenfeld [14]. Shirokov took two angular invariants at each vertex: the Riemannian angle $\Lambda(A)$ between the two sides and the holomorphy inclination $\gamma(A)$. It is seen that

$$\sin(\Lambda(A))\cos(\gamma(A)) = \sin(\Psi(A)).$$

4 Blaschke and Terheggen laws

In $\mathbb{C}^3$, for any vectors x and y, Blaschke and Terheggen defined the product $x \times y$ which is a vector in $\mathbb{C}^3$ with coordinates

$$x_s y_t - x_t y_s, \quad (s,t) \in \{(1,2),(2,0),(0,1)\}$$

and which satisfies

1. $\langle x \times y, \bar{x}\rangle = 0$, $\langle x \times y, \bar{y}\rangle = 0$,
2. $\langle x \times y, u \times v\rangle = \langle x, u\rangle \langle y, v\rangle - \langle x, v\rangle \langle y, u\rangle$,
3. $(x \times y) \times (u \times v) = (xuv)y - (yuv)x$.

Item 2 is Bianchi's identity and (xuv) is the determinant of the three vectors. Now let $([x],[y],[z])$ be a triangle in $\mathbb{C}P^2$ with $\|x\| = \|y\| = \|z\| = 1, a = \arccos(|\langle y, z\rangle|)$, $b = \arccos(|\langle x, z\rangle|)$, and $c = \arccos(|\langle x, y\rangle|)$.

Definition 4.1. The triple $([X],[Y],[Z])$ with

$$X = \frac{\bar{y} \times \bar{z}}{\sin(a)}, \quad Y = \frac{\bar{z} \times \bar{x}}{\sin(b)}, \quad Z = \frac{\bar{x} \times \bar{y}}{\sin(c)}$$

is called the *polar triangle* of the triangle $([x],[y],[z])$.

Now set $|\langle X, Y\rangle| = \cos r$, $|\langle Y, Z\rangle| = \cos p$ and $|\langle Z, X\rangle| = \cos q$. Let us also set $v_1 = u(y,[x])$, $v_2 = u(y,[z])$, $w_1 = u(z,[y])$ and $w_2 = u(z,[x])$. Now consider the complex numbers ω, α and Ω:

$$\langle x, y\rangle \langle y, z\rangle \langle z, x\rangle = \cos(a)\cos(b)\cos(c)e^{i\omega},$$

$$\langle u_2, u_1\rangle \langle v_1, v_2\rangle \langle w_1, w_2\rangle = |\langle u_2, u_1\rangle \langle v_1, v_2\rangle \langle w_1, w_2\rangle| \, e^{i\alpha},$$

$$\langle X, Y\rangle \langle Y, Z\rangle \langle Z, X\rangle = \cos(p)\cos(q)\cos(r)e^{i\Omega}.$$

The invariants ω, Ω, and α are related as follows.

Theorem 4.2. *For any triangle ABC in $\mathbb{C}P^2$,*

1. $\Phi(A) = p$, $\Phi(B) = q$, $\Phi(C) = r$;
2. $\Omega = -\alpha - 2\omega + (2k+1)\pi$, $k \in \mathbb{Z}$.

Proof. A direct computation leads to the following formulae:

$$\langle u_1, u_2 \rangle = \frac{\cos(a)e^{i\omega} - \cos(b)\cos(c)}{\sin(b)\sin(c)}, \tag{14}$$

and

$$\langle Y, Z \rangle = -\frac{\cos(a)e^{-i\omega}\langle u_1, u_2 \rangle}{\langle z, y \rangle}. \tag{15}$$

By a cyclic permutation of a, b, c and x, y, z and X, Y, Z we obtain

$$\langle Z, X \rangle = -\frac{\cos(b)e^{-i\omega}\langle v_2, v_1 \rangle}{\langle x, z \rangle}, \tag{16}$$

$$\langle X, Y \rangle = -\frac{\cos(c)e^{-i\omega}\langle w_2, w_1 \rangle}{\langle y, x \rangle}. \tag{17}$$

Clearly from (15), (16), and (17) we arrive to $\cos\Phi(A) = \cos p$, $\cos\Phi(B) = \cos q$, $\cos\Phi(C) = \cos r$ and

$$\langle X, Y \rangle \langle Y, Z \rangle \langle Z, X \rangle = -e^{-2i\omega}\langle u_1, u_2 \rangle \langle v_2, v_1 \rangle \langle w_2, w_1 \rangle,$$

from which assertion 2 follows. ☐

All the trigonometric laws found by Blaschke and Terheggen, Rozenfeld, Shirokov, Hsiang and Brehm are the same. Of course, the second sine law was not found by Blashke and Terheggen although they knew that a triangle in $\mathbb{C}P^2$ depends on four invariants. The second sine law is self-contained in their investigation and it is derived as follows. With our notation the angles ϕ, χ and ψ of [2] are given by

$$\sin(\phi) = \frac{\cos(a)\sin(\omega)}{\cos(p)\sin(b)\sin(c)} = \frac{\sin(\Psi(A))}{\cos(p)}, \tag{18}$$

$$\sin(\chi) = \frac{\cos(b)\sin(\omega)}{\cos(q)\sin(c)\sin(a)} = \frac{\sin(\Psi(B))}{\cos(q)}, \tag{19}$$

$$\sin(\psi) = \frac{\cos(c)\sin(\omega)}{\cos(r)\sin(b)\sin(a)} = \frac{\sin(\Psi(C))}{\cos(r)}. \tag{20}$$

It follows from the equalitie of the left hand sides of (18), (19), and (20) that

$$\frac{\cos(p)\sin(\phi)}{\sin(2a)} = \frac{\cos(q)\sin(\chi)}{\sin(2b)} = \frac{\cos(r)\sin(\psi)}{\sin(2c)}. \tag{21}$$

Formula (21) is exactly the second sine law.

In [2] or from (14) we find the formula

$$\cos p = \left| \frac{\cos(b)\cos(c)e^{-i\omega} - \cos(a)}{\sin(b)\sin(c)} \right|.$$

By a direct computation, using $p = \Phi(A)$, (2) and (5) we arrive easily to the cosine law (10).

In [2] we find

$$\frac{\sin(p)}{\sin(a)} = \frac{\sin(q)}{\sin(b)} = \frac{\sin(r)}{\sin(c)}.$$

Clearly this is the sine law (8) since $p = \Phi(A)$, $q = \Phi(B)$ and $r = \Phi(C)$. We can then consider that all the laws were found by Blaschke and Terheggen. Item 1 of the above theorem is known in [15] in case of triangles degenerating in real projective planes. On the other hand, Terheggen showed that the work of Study on spherical trigonometry [17] is contained in the trigonometry of the complex projective line [15].

Remark 4.3. Let us turn to Hsiang's laws. Brehm's angular invariants $\Lambda(A)$, $\Phi(A)$, $\Psi(A)$ correspond to the angular invariants $\Lambda(A)$, $\frac{\pi}{2} - \mu(A)$, $\frac{\pi}{2} - \theta(A)$, respectively, which have been defined by Hsiang in [10]. Of course Hsiang derived the same laws of sines but instead of (10) or (13) he obtained a more complicated equivalent equation which we do not give here.

Shirokov took two angular invariants at each vertex: the Riemannian angle $\Lambda(A)$ between the two sides and the holomorphy inclination $\gamma(A)$. By comparing his three laws, we easily see that the holomorphy inclination is linked to Brehm's angles as follows:

$$\sin(\Lambda(A))\cos(\gamma(A)) = \sin(\Psi(A)).$$

The trigonometric laws found by Rozenfeld and Shirokov are the same as (8), (9), (10), or (13) where we replace $\sin(\Psi(A))$ by $\sin(\Lambda(A))\cos(\gamma(A))$ and $\sin(\Phi(A))$ by $\sin(\Lambda(A))\sin(\gamma(A))$ and by a cyclic permutation of the letters A, B, C. This can be found in [14], [16], and [13].

5 Concluding remarks

1. A slight modification of the definitions and computation yields similar results for triangles in complex hyperbolic spaces. The results can be found in [4].

2. It follows from the normal form that each triangle in $\mathbb{K}P^n$ ($n \geq 2$, $\mathbb{K} = \mathbb{C}, \mathbb{H}$) is contained in some totally geodesic $\mathbb{C}P^2$.

3. It is known from Wolf [19] that each isometry of a totally geodesic $\mathbb{C}P^2$ in Cay(P^2) can be extended to the Cayley plane. Thus, it is sufficient to regard triangles of the Cayley plane only in $\mathbb{C}P^2$. The results are also valid for triangles in Cay(H^2).

4. According to Wang [18], $\mathbb{K}P^n$, $\mathbb{K}H^n$, $\mathbb{K} = \mathbb{R}, \mathbb{C}, \mathbb{H}$, Cay$(P^2)$, Cay$(H^2)$ are besides the classical spaces all the two-point homogeneous spaces.

5. Let $([x], [y], [z])$ be a triangle in $\mathbb{C}P^2$ and let S be the surface generated by the geodesic arcs issued from $[x]$ with endpoints on the geodesic arc between $[y]$ and $[z]$. Let I be the integral of the Kähler form on the surface S. Hangan and Masala showed that $\omega = -2I + 2k\pi$, for some integer k, see [8].

6. The result of [8] made more clear a result obtained in [6]. the author constructed regular tetraedra of $\mathbb{C}P^3$ which have symmetry groups isomorphic to the symmetric group S_4. It turns out that a regular tetrahedron with edge $a < \arccos \frac{\sqrt{3}}{3}$ is necessarily contained in a real projective sub-space. Indeed, by regularity and Stokes' theorem, the integral I performed on its faces vanishes and we get from (3) and (6) that $\sin \Psi(A) = 0$. However, in [6], he proved that for any a such that $a \geq \arccos \frac{\sqrt{3}}{3}$ there exists a regular tetrahedron with edge a and with zero shape invariant. In this case, $2I = \frac{\pi}{2} + k\pi$. This means that the quadruple with the geodesic arcs between the vertices and the faces generated by the geodesic arcs issued from a vertex with endpoints on the opposite edge, constitute a simplicial complex of dimension 2 whose homology class of dimension 2 generates the homology group $H_2(\mathbb{C}P^n, \mathbb{Z})$ which is isomorphic to $\mathbb{Z}$ [9].

Trigonometry in general symmetric spaces is still at its infancy.

1. The trigonometry of symmetric spaces of rank 1 was studied by Hsiang in 1989 using integration of the geodesic flow [10]. A different approach using models in projective spaces has been given by Brehm in 1990 [4]. The trigonometry of a complex projective plane was first studied by Blaschke and Terheggen in 1939 [2], Terheggen in 1940 [15] and later pursued in 1957 by Rozenfeld [14] and independently by Sirokov in 1957 [16].

2. Aslaksen in 1989 [1] developed trigonometry for triangles in the Lie group SU(3) using algebraic invariants. Leuzinger in his Ph.D. thesis in 1990 [11] developed trigonometry for arbitrary Riemannian symmetric spaces of non-compact type using the general theory of symmetric spaces.

3. Hangan [9] wrote some trigonometric formulae for Grassmann manifolds.

4. The congruence criteria for triangles in Grassmann manifolds was stated by Masala in his Ph.D. thesis in 1996 using matrices of projections [12].

References

[1] H. Aslaksen, "Laws of trigonometry on SU(3)," *Trans. Amer. Math. Soc.* 317 (1990), no. 1, 127–142.
MR 0961593 Zbl 0695.53037 ☞ 422

[2] W. Blaschke and H. Terheggen, "Trigonometria Hermitiana," *Rend. Sem. Mat. Univ. Roma Ser.* 4 (3) (1939), 153–161.
☞ 413, 418, 420, 421, 422

[3] U. Brehm and B. Et-Taoui, "Congruence criteria for finite subsets of complex projective and complex hyperbolic spaces," *Manuscripta Math.* 96 (1998), no. 1, 81–95.
MR 1624356 Zbl 0908.51001 ☞ 413, 415, 416

[4] U. Brehm, "The shape invariant of triangles and trigonometry in two point homogeneous spaces," *Geom. Dedicata* 33 (1990), no. 1, 59–76.
MR 1042625 Zbl 0695.53038 ☞ 413, 415, 416, 417, 421, 422

[5] J. C. Coolidge, "Hermitian metrics," *Ann. Math* 22 (1921), 11–28.
☞ 418

[6] B. Et-Taoui, "Sur le tétraèdre régulier dans un espace projectif complexe," *Geom. Dedicata* 50 (1994), no. 3, 235–241.
MR 1286376 ☞ 422

[7] B. Et-Taoui, *Sur les systèmes finis de points réguliers dans les espaces projectifs et hyperboliques complexes,* Ph.D. thesis, Université de Mulhouse, Mulhouse, 1994.
☞ 413, 414, 416

[8] T. Hangan and G. Masala, "A geometrical interpretation of the shape invariant for geodesic triangles in complex projective spaces," *Geom. Dedicata* 49 (1994), no. 2, 129–134.
MR 1266268 Zbl 0805.53046 ☞ 422

[9] T. Hangan, "Laws of trigonometry in Grassmann manifolds," *Rend. Sem. Mat. Uni. Pol. Torino* 50 (1992), 91–104.
☞ 422, 423

[10] W.-Y. Hsiang, “On the laws of trigonometries of two-point homogeneous spaces,” *Ann. Global Anal. Geom.* 7 (1989), no. 1, 29–45.
MR 1029843 Zbl 0695.53036 ☞ 414, 421, 422

[11] E. Leuzinger, *On the trigonometry of symmetric spaces,* Ph.D. thesis, ETH, Zürich, 1990.
☞ 422

[12] G. Masala, *Trigonométrie et polyèdres dans les variétés de Grassmann* $G_2(\mathbb{R}^n)$*,* Ph.D. thesis, Université de Mulhouse, Mulhouse, 1996.
☞ 423

[13] R. Ortega and M. Santander, “Trigonometry of complex Hermitian type homogeneous symmetric spaces,” *J. Phys. A* 35 (2002), no. 37, 7877–7917.
MR 1945799 Zbl 1066.53101 ☞ 421

[14] B. A. Rozenfeld, “On the theory of symmetric spaces of rank one,” *Mat. Sb. N.S.* 41(83) (1957), 373–380, in Russian. *Acad. R. P. Romîne. An. Romîno-Soviet. Ser. Mat.-Fiz.* (3) 12 (1958), no. 1 (24), 19–27, in Romanian.
MR 0096271 (Russian) MR 0096272 (Romanian) ☞ 414, 419, 421, 422

[15] H. Terheggen, “Zur analytishen Geometrie auf der Geraden von Hermite als Grenzfall der Geometrie in der Hermitishen Ebene und ihr Zusammenhang mit der gewohnlichen spharichen Trigonometrie,” Jareshbericht Deutschen. Math. Verein 50 (1940), 24–35.
☞ 415, 418, 421, 422

[16] P. A. Shirokov, “On a certain type of symmetric spaces,” *Mat. Sb. N.S.* 41(83) (1957) 361–372, in Russian. *Acad. R. P. Romîne. An. Romîno-Soviet. Ser. Mat.-Fiz.* (3) 12 (1958), no. 1 (24), 5–18, in Romanian.
MR 0096269 (Russian) MR 0096270 (Romanian) ☞ 418, 421, 422

[17] E. Study, *Sphärische Trigonometrie, orthogonale Substitutionen and elliptische Funktionen,* Abhandlungen der math. phys. Klassen der Akademie Leipzig 20, 1893.
☞ 421

[18] H. C. Wang, “Two-point homogeneous spaces,” *Ann. of Math.* (2) 55 (1952). 177–191.
MR 0047345 Zbl 0048.40503 414, 422

[19] J. Wolf, “Elliptic spaces in Grassmann manifolds,” *Illinois J. Math.* 7 (1963), 447–462.
MR 0156295 Zbl 0114.37101 422

Chapter 18

A theorem on equiareal triangles with a fixed base

Victor Pambuccian

Contents

1. Introduction 427
2. Metric planes 429
3. Statement and proof of the theorem in planes with non-Euclidean metric 431
References 436

1 Introduction

Inspired by Lexell's theorem (see [9] and [10]) in spherical geometry, which states that for two fixed points A and C on a sphere, the locus of the third vertex B of a spherical triangle ABC of constant area is a small circle, N. A'Campo and A. Papadopoulos [1, Theorem 5.11] have shown that, in plane hyperbolic geometry, the locus of the third vertex B, which lies on the same side of AC, is a hypercycle. Nevertheless, as shown by A. Papadopoulos and W. Su [15, Proposition 4.4], the midpoints of the segments BA and BC lie on a line as B moves along that hypercycle. Thus, although the locus of B bears no resemblance to its Euclidean counterpart, the locus of the midpoints of the segments BA and BC is identical to that in the Euclidean case: a line. This raises the question whether the result could be stated and proved inside absolute geometry. A first candidate would be Hilbert's plane absolute geometry $\mathcal{A}$, whose axioms are the plane axioms of groups I, II, and III of Hilbert's *Grundlagen der Geometrie* (Foundations of Geometry) [6], or equivalently axioms A1–A9 in [16]. Its models will be referred to as *Hilbert planes* (see also [5]). The notion of area in Hilbert planes has been investigated in [4]. According to its results, two triangles ABC and $AB'C$, with B and B' on the same side of AC, have the same area in the case of the Euclidean metric (that is, if there is a rectangle in the Hilbert plane) if and only if the altitude from B to AC and the altitude from B' to AC are congruent. In the case with non-Euclidean metric (that is, if there is no rectangle in the Hilbert plane) triangles ABC and $AB'C$, with B and B' on the same side of AC, have the same area if and only if the sum of the angles is the same in both triangles. At first sight it seems that, for the non-Euclidean case, the axioms for Hilbert planes are needed to even state the result, for one needs order, that is the notion of betweenness, to state that the points B and B' lie on the same side of AC, and one needs the ability to transport angles so that one can meaningfully refer to the sum of three angles that do not have the same vertex. So, it seems that order and free mobility (the fact that, given two distinct points A and B, there is a rigid motion (a one-to-one and onto mapping preserving congruence and betweenness) mapping A onto B, and given two distinct lines g and h, there is a rigid motion mapping g onto h) are required to even state the problem in its current form.

While this is true, one can operate a small change in the statement of the problem and state it in a more general geometry than $\mathcal{A}$. First, let us see how we can express area equality in the case of a non-Euclidean metric. Given triangle ABC, let U denote the midpoint of BC, let V denote the midpoint of AC, and let W denote the midpoint of AB. Let R denote the reflection of V in U, and let Q denote the reflection of V in W. Since $\widehat{VCU} \equiv \widehat{RBU}$ and $\widehat{VAW} \equiv \widehat{QBW}$, the angle $\widehat{QBR}$ represents the sum of the angles of triangle ABC. Since $BR \equiv BQ \equiv VC$, and VC is constant, given that A and C are fixed points, the isosceles triangle BQR has its two congruent sides of fixed length, and the angle between two sides is congruent to the sum of the angles of triangle ABC. The sum of the angles of triangles ABC and $AB'C$ are thus the same if and only if $QR \equiv Q'R'$, where by Q' and R' we have denoted the points obtained in the manner Q and R were, only starting with triangle $AB'C$. It thus turns out that only the assumption that any segment has a midpoint is needed, not the entire assumption of free mobility. We also notice that order is not necessary, for there are two lines on which the midpoints of the sides BA and BC can lie, and one is the reflection of the other in the line AC. So we need a geometry in which we can express segment congruence, and in which any pair of points has a unique midpoint.

The congruence core of plane absolute geometry was investigated in a series of papers by J. Hjelmslev [7] and [8], and reached a particularly simple form in the axiomatics presented by F. Bachmann [2] for structures called *metric planes*. We will see that our theorem can be expressed and proved in the theory of non-elliptic metric planes in which every segment has a midpoint. Why “non-elliptic”? Metric planes are called *elliptic* if there are three line reflections in them whose composition is the identity. This amounts to the existence of points which have several perpendiculars to a given line. Under those circumstances the uniqueness of midpoints is no longer available and the whole edifice of our reduction of the area problem to one about reflections crumbles.

2 Metric planes

Metric planes, as presented in [2], are defined in the language of groups. The idea behind the axiomatization is to focus on line reflections and on their properties when composed with other line reflections. One notices that if a and b are line reflections, and it so happens that $ab = ba$, then the lines defined by the reflections a and b must be perpendicular, and conversely. Since perpendicular lines intersect, and the composition of the reflections in them amounts to the reflection in their point of intersection, we can choose to refer to the product ab, in case $ab = ba$, as to a *point reflection*. Another familiar geometric notion, that of point-line incidence can now be expressed in the language of line reflections. The point P determined by the point reflection ab, where a and b denote line reflections with $ab = ba$, is said to be incident with the line determined by the line reflection g if $Pg = gP$.

The axiom system consists of a *fundamental assumption*, that G be a group (written multiplicatively) generated by an invariant set S of involutory elements, as well as five axioms (see below). The elements of S will be denoted by lowercase Latin letters, and will be referred to as *line reflections* (or simply *lines*). Involutory products of two elements in S will be denoted by uppercase Latin letters, and will be referred to as *point reflections* (or simply *points*). For any two elements α and β in G, $\alpha \mid \beta$ denotes the fact that $\alpha \cdot \beta$ is involutory. The notation $\alpha_1, \ldots \alpha_n \mid \beta_1, \ldots \beta_m$ stands for the conjunction of all $\alpha_i \mid \beta_j$ with $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m\}$. The relation $P \mid g$ may be read as “point P is incident with line g,” and $g \mid h$ as “line g is orthogonal to line h.” The axioms for metric planes state that

A1. *For all points P and Q there is a line g such that $P, Q \mid g$.*

A2. *If, for points P and Q and lines g and h, we have $P, Q \mid g, h$, then $P = Q$ or $g = h$.*

A3. *If the lines a, b, and c and the point P are such that $a, b, c \mid P$, then there is a line d such that $abc = d$.*

A4. *If the lines a, b, c, and g are such that $a, b, c \mid g$, then there is a line d such that $abc = d$.*

A5. *There are lines g, h, j such that $g \mid h$, but none of $j \mid g$, $j \mid h$, or $j \mid gh$ holds.*

By **A1** and **A2** there is a unique line joining two distinct points P and Q. We will denote it by $\langle P, Q\rangle$. It is, in general, not true that a midpoint exists for every pair of points. In this context, we say that M is a midpoint for A and B if $A^M = B$. Here we have denoted by α^β the element $\beta^{-1}\alpha\beta$. Nor is it in general true that a midpoint, should it exist, is unique. In fact, to ensure uniqueness (see [2, pp. 53-54]), it is necessary and sufficient to assume that the metric plane is not elliptic. This means that it satisfies the axiom

∼P. *For all a, b, c, we have $abc \neq 1$.*

An alternate requirement which is equivalent to $\sim \boldsymbol{P}$, is that, for any line l and any point P not incident with l, there exists a unique line g incident with P and perpendicular to l. This requirement is obviously false in the case of plane elliptic geometry, in which, for every line there is a point from which one can draw infinitely many perpendiculars to that line (that that there are infinitely many perpendiculars from a *pole* of a line to that line follows from [2, Satz 20, §6,12, p. 121], which states that there are no finite elliptic planes, so every line must be incident with infinitely many points). Double elliptic geometry is already excluded by the axioms for metric planes, as there is a unique line joining two distinct points in metric planes.

Since the theorem we want to prove depends both for the definition of the area and for its conclusion on the uniqueness of midpoints, we will be interested only in non-elliptic metric planes. Since we also need the unrestricted existence of midpoints, we also assume the axiom

M. *For all A, B, there exists M such that $A^M = B$.*

Metric planes can have a Euclidean metric, which means that there exists a rectangle, i.e.

R. *There are a, b, c, d such that $a, b \mid c, d$ and $a \neq b$, $c \neq d$.*

or a non-Euclidean metric, which means that there is no rectangle, i.e.

∼R. *If $a, b \mid c, d$, then $a = b$ or $c = d$.*

For metric planes satisfying **R**, it can be easily seen that the midlines $\langle U.W\rangle$ and $\langle U', W'\rangle$ of two triangles ABC and $AB'C$ having the same area—where U and W are the midpoints of (B, C) and (B, A), and U' and W' are the midpoints of (B', C) and (B', A)—are either identical or one is the reflection of the other in $\langle A, C\rangle$, given that, in the presence of **R**, the area equality of ABC and $AB'C$ is equivalent to the statement that $B = B'$ or $\langle B, B' \rangle \parallel \langle A, C \rangle$, or the midpoint of BB' lies on the line $\langle A, C \rangle$ (here $\parallel$ stands for the notion of Euclidean parallelism, with $g \parallel h$ being defined to stand for "for all l, $l \mid g$ if and only if $l \mid h$").

The interesting case is thus that in which $\sim$**R** holds.

Non-elliptic metric planes have been axiomatized in various other languages in [11], [17], [12], [13], and [14]. The Hilbert planes are precisely the ordered metric planes with free mobility. Although there are Hilbert planes in which the sum of the angles of a triangle exceeds two right angles, as was discovered by Max Dehn in [3], these are non-elliptic in the sense that the perpendicular from a given point to a given line is always unique. The phrasing and proof of our theorem inside non-elliptic metric planes is thus more general than any version thereof inside Hilbert's plane absolute geometry.

3 Statement and proof of the theorem in planes with non-Euclidean metric

The theorem we will now prove inside the theory axiomatized by the fundamental assumption, **A1**–**A4**, **M**, $\sim$**P**, and $\sim$**R** can be stated as follows (we do not need **A5**, for, in the absence of the elements it stipulates to exist, the hypothesis of our theorem, that there are three non-collinear points A, B, C, is false, and thus the theorem is true):

Theorem 3.1. *If* $A \neq C$, $A, C \mid b$, $B \nmid b$, $C^W = B$, $B^U = A$, $A^V = C$, $U, W \mid s$, $Q = V^W$, $R = V^U$, $B' \nmid b$, $C^{W'} = B'$, $B'^{U'} = A$, $U', W' \mid s'$, $Q' = V^{W'}$, $R' = V^{U'}$, $Q', Q^{UU'} \mid e$, $R' \mid f$, $e \mid f$, *and* $Q^{UU'f} = Q'$, *then* $s = s'$ *or* $s' = s^b$.

Proof. Before starting the proof, we should explain why the statement above reflects the intentions of the theorem we want to prove. As mentioned in the introduction, we have reduced the original problem to showing that, if R is the reflection of V in U, Q is the reflection of V in W, and similarly for R' and Q', and $(R, Q) \equiv (R', Q')$, then the lines $s = \langle U, W \rangle$ and $s' = \langle U', W' \rangle$ either coincide or one is the reflection of the other in the line $\langle A, C \rangle$. To make sense of the segment congruence relation, we observe that we can define it in metric planes by stipulating that $(A, B) \equiv (C, D)$ if and only if there exists an element $\alpha \in G$ such that $C = A^\alpha$ and $D = B^\alpha$.

We now turn to the reason why the conditions $Q', Q^{UU'} \mid e$, $R' \mid f$, $e \mid f$, and $Q^{UU'f} = Q'$ amount to saying that $(R, Q) \equiv (R', Q')$. Since $R^U = V$ and $V^{U'} = R'$, we have that $R^{UU'} = R'$. By our definition of segment congruence we thus conclude that $(R, Q) \equiv (R', Q^{UU'})$. However, we also have $(R', Q^{UU'}) \equiv (R', Q')$, for if $Q^{UU'} = Q'$, there is nothing to prove (and f will have to be chosen to be $\langle R', Q' \rangle$ and e the perpendicular raised in Q' on $\langle R', Q' \rangle$), and if $Q^{UU'} \neq Q'$, then f is the unique perpendicular to $\langle Q^{UU'}, Q' \rangle$ passing through R', and $(R', Q^{UU'}) \equiv (R'^{f}, Q^{UU'f})$, which, given that $R'^{f} = R'$ and $Q^{UU'f} = Q'$, amounts to $(R', Q^{UU'}) \equiv (R', Q')$.

We are now ready for the proof. Let v be such that $V \mid v$ and $v \mid s$, and let v' be such that $B \mid v'$ and $v' \mid s$ (these perpendiculars dropped from a given point to a given line exist by [2, Satz 2 of §3,4, p. 37]). According to [2, Satz 2 of §4,1, p. 57], we have $Uv'W = v$ and $C^{v} = A$. Let M be the midpoint of vs and $v's$, thus a point for which $v^{M} = v'$. We now have

$$(v^{W})^{M} = v'^{U}, \tag{1}$$

$$(v^{U})^{M} = v'^{W}. \tag{2}$$

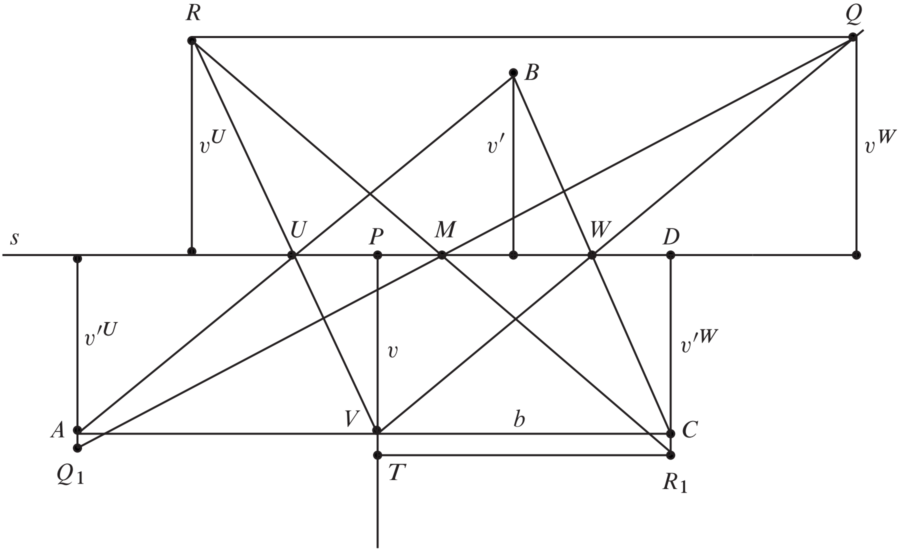


Figure 1. The triangle ABC with all relevant points and lines

To prove (1), first notice that $v^{M} = MvM$ and $v^{M} = v' = UvW$, so $MvM = UvW = WvU$ (given that UvW is a line, thus involutory), and thus (*) $MvMU = Wv$. Using (*), we get $(v^{W})^{M} = v^{WM} = M(Wv)WM = M(MvMU)WM = vMUWM$. On the other hand, $v'^{U} = Uv'U = UUvWU = vWU$. To prove (1), we thus need to show that $vMUWM = vWU$, i.e., that $MUWM = WU$, which amounts to $(MUW)^2 = 1$. This holds, given that the points M, U, and W are incident with the line s, and the compositions of reflections in three collinear points is a reflection in a point, thus involutory (see [2, Satz 24b of §3,9, p. 49]). To prove (2), first notice that $v'^W = (WvU)^W = WWvUW = vUW$. By (*), we have $Uv = MvMW$, so $v^{UM} = M(Uv)UM = M(MvMW)UM = vMWUM$. To prove (2), we thus need to show that $vMWUM = vUW$, i.e., that $MWUM = UW$, which amounts to $(MWU)^2 = 1$, which holds for the reason already mentioned in the proof of (1).

It is easily checked that if $\alpha \mid \beta$, then $\alpha^\sigma \mid \beta^\sigma$. Given that $V \mid v$, we have $V^W \mid v^W$, and so $(V^W)^M \mid (v^W)^M$, i.e., by (1), $(V^W)^M \mid v'^U$. Also by $V \mid v$, we have $V^U \mid v^U$, and so $(V^U)^M \mid (v^U)^M$, i.e., by (2), $(V^U)^M \mid v'^W$.

Note that, again by [2, Satz 2 of §4,1, p. 57], v' is the perpendicular bisector of segment (Q, R), and thus, given that $v'^M = v$ and that orthogonality and midpoints are preserved under rigid motions, we have that v is the perpendicular bisector of (Q_1, R_1), where by Q_1 and R_1 we have denoted Q^M and R^M.

Let $P = sv$, $D = v'^W s$, let g denote the line $\langle Q_1, R_1\rangle$ joining Q_1 and R_1, and $T = vg$. We will now focus on the following configuration of points and lines: $V, v, s, b, P, C, v'^W, D$ R_1, g, T. The elements of our configuration are in the following relations: $V \mid v, b$; $D \mid s.v'^W$; $C \mid b, v'^W$; $P \mid s, v$; $R_1 \mid g, v'^W, T \mid v, g$; $v \mid b, s, g$; $v'^W \mid s$; $(V, P) \equiv (R_1, D)$. The latter holds since $V^{UM} = R_1$ and $P^{UM} = D$. We also have that (R_1, T) is half of (Q_1, R_1). Following the same steps, with B replaced by B', U by U', and W by W', we construct the points R_1' and T'.

Since we could not see how to prove this theorem by purely reflection-geometric considerations, we will use the fact that all non-elliptic metric planes with a non-Euclidean metric can be embedded in a projective-metric plane over some field of characteristic $\neq 2$, containing its origin, such that the image of the point set of the non-elliptic metric plane does not intersect the line at infinity (the *polar* of the origin). This major result and the formulas used below are the subject of [2, §6, pp. 93–127, §8,1, pp. 141–144]. In the embedded non-elliptic metric planes, the points can be written as pairs (a, b), with $a, b \in F$, and the lines can be written in homogeneous coordinates as triples $[u, v, w]$, not all of u, v, w being allowed to be 0, two such triples representing the same line if they can be obtained from each other by multiplying each coordinate with the same non-zero element of F. The incidence of (a, b) with $[u, v, w]$ (to be denoted by $(a, b) \mid [u, v, w]$) is given by $au + bv + w = 0$, whereas the orthogonality of $[u, v, w]$ and $[u', v', w']$ (to be denoted by $[u, v, w] \mid [u', v', w']$) is given by $\lambda uu' + \mu vv' + \nu ww' = 0$, where λ, μ, ν are fixed elements of F, all different from 0. The congruence of two segments $((a, b), (c, d))$ and $((a', b'), (c', d'))$

is given by

$$\frac{F((a,b),(c,d))^2}{Q((a,b))Q((c,d))} = \frac{F((a',b'),(c',d'))^2}{Q((a',b'))Q((c',d'))},$$

where

$$F((x,y),(x',y')) = \mu\nu xx' + \lambda\nu yy' + \lambda\mu$$

and

$$Q((x,y)) = F((x,y),(x,y)).$$

We can embed the metric plane in such a way that V is mapped onto $(0,0)$, b is mapped onto $[0,1,0]$ and v onto $[1,0,0]$. Then, for some $a, x \in F$, $C = (a,0)$, $P = (0,x)$, $s = [0,1,-x]$. Since v'^W is incident with $(a,0)$ and orthogonal to $[0,1,-x]$, if we denote v'^W by $[m,n,p]$, we have $am + p = 0$ and $\mu n - \nu xp = 0$. Thus $v'^W = [\mu, -\nu xa, -a\mu]$. Since it lies on s, $D = (\delta, x)$, for some $\delta \in F$. Since D is on v'^W, we have $(\delta, x) \mid [\mu, -\nu xa, -a\mu]$, and so $\delta = a\bigl(1 + \frac{\nu}{\mu}x^2\bigr)$, i.e., $D = \bigl(a\bigl(1 + \frac{\nu}{\mu}x^2\bigr), x\bigr)$. Regarding the coordinates (α, β) of R_1, we have the following two relations that R_1 is in

(i) $(\alpha,\beta) \mid [\mu, -\nu xa, -a\mu]$

and

(ii) $((0,0),(0,x)) \equiv \Bigl(\Bigl(a\Bigl(1 + \dfrac{\nu}{\mu}x^2\Bigr), x\Bigr), (\alpha,\beta)\Bigr)$.

From (i) we get that $\alpha = a\bigl(\frac{\nu}{\mu}\beta x + 1\bigr)$. Since

$$\frac{F((0,0),(0,x))^2}{Q((0,0))Q((0,x))} = \frac{\mu}{\nu x^2 + \mu}$$

and

$$\frac{F\bigl(\bigl(a(1 + \frac{\nu}{\mu}x^2), x\bigr), \bigl(a\bigl(\frac{\nu}{\mu}\beta x + 1\bigr), \beta\bigr)\bigr)^2}{Q\bigl(\bigl(a\bigl(1 + \frac{\nu}{\mu}x^2\bigr), x\bigr)\bigr) Q((a(\frac{\nu}{\mu}\beta x + 1), \beta))}$$
$$= \frac{\mu\bigl(\frac{\nu}{\mu}\beta x + 1\bigr)^2\bigl(\nu a^2\bigl(1 + \frac{\nu}{\mu}x^2\bigr) + \lambda\bigr)}{\bigl(1 + \frac{\nu}{\mu}x^2\bigr)\bigl(\mu\nu a^2\bigl(\frac{\nu}{\mu}\beta x + 1\bigr)^2 + \lambda\nu\beta^2 + \lambda\mu\bigr)},$$

(ii) gives

$$\mu\nu a^2\Bigl(\frac{\nu}{\mu}\beta x + 1\Bigr)^2 + \lambda\nu\beta^2 + \lambda\mu = \mu\Bigl(\frac{\nu}{\mu}\beta x + 1\Bigr)^2\Bigl(\nu a^2\Bigl(1 + \frac{\nu}{\mu}x^2\Bigr) + \lambda\Bigr). \quad (3)$$

Our hypothesis states, as noted at the start of the proof, that $(Q, R) \equiv (Q', R')$. Since (R_1, T) is half of (Q, R), and (R'_1, T') is half of (Q', R'), our hypothesis implies that $(R_1, T) \equiv (R'_1, T')$. With $R'_1 = (\alpha', \beta')$, we thus have

$$\frac{F((0,\beta),(\alpha,\beta))^2}{Q((0,\beta))Q((\alpha,\beta))} = \frac{F((0,\beta),(\alpha',\beta'))^2}{Q((0,\beta'))Q((\alpha',\beta'))}. \tag{4}$$

Applying (3) to

$$\frac{F((0,\beta),(\alpha,\beta))^2}{Q((0,\beta))Q((\alpha,\beta))} = \frac{\lambda\nu\beta^2 + \lambda\mu}{\mu\nu a^2\left(\frac{\nu}{\mu}\beta x + 1\right)^2 + \lambda\nu\beta^2 + \lambda\mu},$$

we get

$$\frac{F((0,\beta),(\alpha,\beta))^2}{Q((0,\beta))Q((\alpha,\beta))} = \frac{\lambda\nu\beta^2 + \lambda\mu}{\mu\left(\frac{\nu}{\mu}\beta x + 1\right)^2\left(\nu a^2\left(1 + \frac{\nu}{\mu}x^2\right) + \lambda\right)}. \tag{5}$$

Applying (3) again, we get

$$\lambda\nu\beta^2 + \lambda\mu = \left(\frac{\nu}{\mu}\beta x + 1\right)^2(\nu^2 a^2 x^2 + \lambda\mu)$$

and thus

$$\frac{\lambda\nu\beta^2 + \lambda\mu}{\mu\left(\frac{\nu}{\mu}\beta x + 1\right)^2\left(\nu a^2\left(1 + \frac{\nu}{\mu}x^2\right) + \lambda\right)} = \frac{\nu^2 a^2 x^2 + \mu\lambda}{\mu\left(\nu a^2\left(1 + \frac{\nu}{\mu}x^2\right) + \lambda\right)}. \tag{6}$$

Since the value of a is fixed, as it is the non-zero coordinate of the fixed point C, equations (4)–(6) give

$$\frac{\nu^2 a^2 x^2 + \mu\lambda}{\mu\left(\nu a^2\left(1 + \frac{\nu}{\mu}x^2\right) + \lambda\right)} = \frac{\nu^2 a^2 x'^2 + \mu\lambda}{\mu\left(\nu a^2\left(1 + \frac{\nu}{\mu}x'^2\right) + \lambda\right)}, \tag{7}$$

which implies $x^2 = x'^2$, and thus $x = x'$ or $x = -x'$. This proves our theorem. □

We conclude that, with the corresponding definitions for area equality, it is true in the theory axiomatized by the fundamental assumption and {**A1**–**A4**, **M**, ∼**P**} that

Theorem 3.2. *If ABC and $AB'C$ have the same area and $\langle U, W\rangle$ and $\langle U', W'\rangle$ are their respective midlines, then they either coincide or one is the reflection of the other in $\langle A, C\rangle$.*

Acknowledgement The author thanks Vincent Alberge for a very close reading of the manuscript and for very useful suggestions for improvement.

References

[1] N. A'Campo and A. Papadopoulos, "Notes on non-Euclidean geometry," in *Strasbourg master class on geometry* (A. Papadopoulos, ed.), IRMA Lectures in Mathematics and Theoretical Physics 18, European Mathematical Society (EMS), Zürich, 2012, 1–182.
MR 2931886 Zbl 1253.51001 ☞ 427

[2] F. Bachmann, *Aufbau der Geometrie aus dem Spiegelungsbegriff,* second enlarged edition, Die Grundlehren der mathematischen Wissenschaften 96, Springer-Verlag, Berlin etc., 1973.
MR 0346643 Zbl 0085.14502 ☞ 428, 429, 430, 432, 433

[3] M. Dehn, "Die Legendre'schen Sätze über die Winkelsumme im Dreieck," *Math. Ann.* 53 (1900), no. 3, 404–439.
MR 1511095 JFM 31.0471.01 ☞ 431

[4] A. Finzel, "Die Lehre vom Flächeninhalt in der allgemeinen Geometrie," *Math. Ann.* 72 (1912), no. 2, 262–284.
MR 1511697 JFM 43.0565.03 ☞ 427

[5] G. Hessenberg and J. Diller, *Grundlagen der Geometrie,* second edition, Walter de Gruyter, Berlin, 1967.
Zbl 0158.39001 ☞ 427

[6] D. Hilbert, *Foundations of geometry,* second edition, translated from the tenth German edition by L. Unger, Open Court, LaSalle, IL, 1971.
MR 0275262 ☞ 427

[7] J. Hjelmslev, "Neue Begründung der ebenen Geometrie," *Math. Ann.* 64 (1907), no. 4, 449–474.
MR 1511450 JFM 38.0504.01 ☞ 428

[8] J. Hjelmslev, "Einleitung in die allgemeine Kongruenzlehre I., II.," *Mat.-Fys. Medd. K. Dan. Vidensk. Selsk.* 8 (1929), no. 11, 36pp.; ibid. 10 (1929), no. 1, 28pp.
JFM 55.0957.02 ☞ 428

[9] A. J. Lexell, "Solutio problematis geometrici ex doctrina sphaericorum," *Acta Academiae Scientarum Imperialis Petropolotinae* 1 (1781), 112–126.
☞ 427

[10] H. Maehara and H. Martini, "On Lexell's theorem," *Amer. Math. Monthly* 124 (2017), no. 4, 337–344.
MR 3626255 Zbl 1391.51017 ☞ 427

[11] H. Müller, *Zur Begrüundung der ebenen absoluten Geometrie aus Bewegungsaxiomen,* Dissertation, Technische Hochschule München, 1966.
☞ 431

[12] V. Pambuccian, "Constructive axiomatization of non-elliptic metric planes," *Bull. Polish Acad. Sci. Math.* 51 (2003), no. 1, 49–57.
MR 1990959 Zbl 1036.03051 ☞ 431

[13] V. Pambuccian, "Groups and plane geometry," *Studia Logica* 81 (2005), no. 3, 387–398.
MR 2195502 Zbl 1086.03007 ☞ 431

[14] V. Pambuccian, "Orthogonality as single primitive notion for metric planes," with an appendix by Horst and Rolf Struve, *Beiträge Algebra Geom.* 48 (2007), no. 2, 399–409.
MR 2364798 Zbl 1161.51004 ☞ 431

[15] A. Papadopoulos and W. Su, "On hyperbolic analogues of some classical theorems in spherical geometry," in *Hyperbolic geometry and geometric group theory* (K. Fujiwara, K. Ohshika, and S. Kojima, eds.), Advanced Studies of Pure Mathematics 73, Mathematical Society of Japan, Tokyo, 2017, 225–253.
MR 3728500 ☞ 427

[16] W. Schwabhäuser, W. Szmielew, and A. Tarski, *Metamathematische Methoden in der Geometrie,* Springer, Berlin, 1983; re-issued by ISHI Press International, Bronx, N.Y., 2011.
MR 0731370 Zbl 0564.51001 ☞ 427

[17] K. Sörensen, "Ebenen mit Kongruenz," *J. Geom.* 22 (1984), no. 1, 15–30.
MR 0756947 Zbl 0537.51019 ☞ 431

List of contributors

Annette A'Campo-Neuen, 237*, 253
Departement Mathematik und Informatik, Universität Basel, Spielgasse 1,
CH-4051 Basel, Switzerland

annette.acampo@unibas.ch

Norbert A'Campo, 3
Departement Mathematik und Informatik, Universität Basel, Spielgasse 1,
CH-4051 Basel, Switzerland

norbert.acampo@unibas.ch

Nikolay Abrosimov, 151
Sobolev Institute of Mathematics, pr. Koptuga, 4,
and Novosibirsk State University, Pirogova st., 2, 630090, Novosibirsk,
Russian Federation

abrosimov@math.nsc.ru

Vincent Alberge, v, vii, 47
Department of Mathematics, Fordham University, 441 East Fordham Road, Bronx,
NY 10451, USA

valberge@fordham.edu

Charalampos Charitos, 125, 135
Agricultural University Athens, Iera Odos 75, 11855 Athens, Greece

bakis@aua.gr

Boumediene Et-Taoui, 413
LMIA, Université de Haute Alsace, 4 rue des frères Lumière, 68093 Mulhouse Cedex,
France

boumediene.ettaoui@uha.fr

---

* Translator.

François Fillastre, 321

Laboratoire AGM, Université de Cergy-Pontoise, 2 avenue Adolphe Chauvin,
95302 Cergy-Pontoise Cedex, France

`francois.fillastre@u-cergy.fr`

Elena Frenkel, 27, 47

Institut de Recherche Mathématique Avancée, Université de Strasbourg et CNRS,
7 rue René Descartes, 67084 Strasbourg Cedex, France

`frenkel@math.unistra.fr`

Ivan Izmestiev, 191, 263

Département de mathématiques, Université de Fribourg, Chemin du Musée 23,
CH-1700 Fribourg, Switzerland,

`ivan.izmestiev@unifr.ch`

Son Lam Ho, 113

Département de mathématiques, Université de Sherbrooke,
2500 boulevard de l'Universitée, Sherbrooke, QC J1K 2R1, Canada

`lam.son.ho@usherbrooke.ca`

Alexander Mednykh, 151

Sobolev Institute of Mathematics, pr. Koptuga, 4,
and Novosibirsk State University, Pirogova st., 2, 630090, Novosibirsk,
Russian Federation

`mednykh@math.nsc.ru`

Victor Pambuccian, 427

School of Mathematical and Natural Sciences, Arizona State University,
West campus, P. O. Box 37100, Phoenix, AZ 85069-7100, USA

`pamb@asu.edu`

Ioannis Papadoperakis, 125

Agricultural University Athens, Iera Odos 75, 11855 Athens, Greece

`papadoperakis@aua.gr`

Athanase Papadopoulos, V, VII, XXI*, 3, 253

Institut de Recherche Mathématique Avancée, Université de Strasbourg et CNRS,
7 rue René Descartes, 67084 Strasbourg Cedex, France

`athanase.papadopoulos@math.unistra.fr`

Himalaya Senapati, 57, 67, 81

Department of Physics, Chennai Mathematical Institute, Chennai,
H1, SIPCOT IT Park, Siruseri Kelambakkam 603103, India

`himalay@cmi.ac.in`

Andrea Seppi, 321

Département de mathématiques, Université du Luxembourg, Maison du nombre,
6 avenue de la Fonte, L-4364 Esch-sur-Alzette, Luxembourg

`andrea.seppi@uni.lu`

Dmitriy Slutskiy, 93

Laboratoire AGM, Université de Cergy-Pontoise, 2 avenue Adolphe Chauvin,
95302 Cergy-Pontoise Cedex, France

`dmitry.slutsky@gmail.com`

Weixu Su, 27

Department of Mathematics, Fudan University, 200433 Shanghai, China

`suwx@fudan.edu.cn`

---

* Translator.

Index

absolute, 266, 327
absolute cone, 241, 254
absolute geometry, 428
absolute polarity, 267
affine field, 387
angle deficit, 5, 13, 17, 43
angle excess, 5, 13, 43, 114, 121
angle of parallelism, 29
anti-de Sitter space, 333
anti-isometry, 324
area, 3, 7, 8, 13, 15, 17, 18, 21, 28, 152
  – Euler formula, 3, 8, 18, 28, 152
    – hyperbolic analogue, 18, 37
  – in a Hilbert plane, 427
  – maximal, 48
area-preserving geographical map, 137
  – distorsion, 138
axioms for a metric plane, 429

Bachmann axioms for metric planes, 428
Bilinski formula for a cyclic hyperbolic quadrilateral, 157
Blaschke, Wilhelm, 415, 418
blow-up
  – of a line, 364
  – of a plane, 369
  – of a point, 362
Bolyai, János, 94
Brahmagupta formula, 154
Brahmagupta formula for bicentric hyperbolic quadrilateral, 156
Bretschneider formula, 159
  – spherical, 163
Busemann curvature, 3, 20, 21, 81
Busemann, Herbert, 9, 20, 259
Busemann-Petty problem, 9

Cagnoli theorem, 152
Casey theorem, 164
Castillon problem, 68
Castillon, Jean de, 67
Catalan, Eugène-Charles, 68
Cauchy development, 359
Cavalieri principle, 8
Cavalieri, Bonaventura Francesco, 7, 8
Cayley, Arthur, 258
Cayley–Klein metric, 339
Chasles theorem, 272
co-Euclidean space, 346
co-Minkowski space, 354
Codazzi equation, 389
complex projective space
  – trigonometry, 413
cone
  – convex, 333
  – dual, 334
confocal conics, 270, 303
conic
  – dual, 265
  – hyperbolic, 294
  – polar, 268
  – spherical, 276
conjecture
  – Seidel, 171
    – strengthened, 171
conjugate lines, 267

conjugate points, 267
connection (Levi–Civita), 373
constructibility
  – spherical geometry, 72
convex
  – body, 343
  – cone, 333
  – set, 335
cross-ratio, 328
curvature (in the sense of Busemann), 3, 20, 21, 81
cyclic polygon, 151

Darboux–Sauer correspodence, 192
de Sitter plane, 294
deformation
  – infinitesimal isometric, 400
Dehn, Max, 431
de Sitter space, 332
dissection, 43
distorsion
  – geographical map, 138
distortion
  – area-preserving geographical map, 138
dual cone, 334

elliptic space, 330
equation
  – Codazzi, 389
  – Gauss, 389
equidistant coordinates, 30
Euclidean space, 342
Euler
  – geographical map, 136
  – perfect map, 125
Euler theorem, 3, 8, 18, 28, 152
Euler, Leonhard, 8, 125, 135
exterior hyperbolic plane, 238, 241, 243, 244, 248, 250, 254
external tangent of two circles, 164

formula
  – Euler (area), 3, 8, 18, 28, 152
  – Laguerre, 358
  – Weyl, 382
Fréchet, Maurice, 257
framework, 191, 194
  – isometric deformation, 195
    – examples, 201–203
    – infinitesimal, 195
  – load, *see* load on a framework
  – polyhedral lift
    – Euclidean radial, 217
    – Euclidean vertical, 217
    – hyperbolic, 226
    – spherical strong, 222
    – spherical weak, 222
  – reciprocal diagram
    – Euclidean, 217
    – hyperbolic, 225
    – spherical strong, 221
    – spherical weak, 221
  – rigidity, *see* rigidity of a framework
  – stress, 206
    – projective, 210
fundamental form
  – first, 388
  – second, 388
  – third, 390
fundamental theorem of immersed surfaces, 390
Fuss, Nicolas, 67

Galilean plane, 371
Gauss equation, 389
Gauss, Carl Friedrich, 9, 13, 15, 94
Gauss–Bonnet theorem, 113, 119, 120
geodesically complete, 326

geographical map, 136, 138
  – distorsion, 138
  – Euler, 136
  – Lambert, 137
geometric transition, 359
geometry (as (X, G) structure), 341
Girard theorem, 3–5, 7, 8, 13, 43
  – hyperbolic analogue, 43
  – proof by Lambert, 13
Girard, Albert, 7, 8
Gram matrix of a tetrahedron, 171

half-pipe geometry, 322
Hausdorff, Felix, 259
Hermitian trigonometry, 413
Heron formula, 21
  – hyperbolic, 21, 152
  – spherical, 21
Hilbert
  – absolute geometry, 427
  – axioms, 427
  – plane, 427
Hilbert metric, 339
Hilbert, David, 259
homography, 325
Hurwitz, Adolf, 259
hyperbolic
  – Brahmagupta formula, 155
  – Bretschneider formula, 160
  – Casey theorem, 164
  – Ptolemy theorem, 153
hyperbolic conic, 294
  – absolute point, 295
  – absolute tangent, 295
  – center, 298
  – classification, 295
  – director point, 300
  – directrix, 300
  – focal line, 297
  – focus, 297
hyperbolic Heron formula, 152
hyperbolic space, 331

Ibn al-Haytham–Lambert quadrilateral, 94
ideal tetrahedron, 169
infinitesimal isometric deformation, 400
infinitesimal Pogorelov map, 192, 211
  – affine, 212
  – hyperbolic, 214
  – projective, 212
  – spherical, 214
infinitesimal rigidity, 400
involutory element, 429
isotropic
  – cone, 324
  – line, 341
Ivory lemma
  – hyperbolic, 304
  – spherical, 282

Jacobi theorem, 176

Killing field, 383
Klein, Felix, 259, 260

Lagrange, Joseph-Louis, 7, 67, 136
Laguerre formula, 358
Lambert
  – proof of Girard theorem, 13
Lambert cylindrical equal area projection, 137
Lambert quadrilateral, 94
Lambert, Johann Heinrich, 13, 59, 93, 137
Levi–Civita connection, 373
Lexell problem, 28, 42–44
Lhuillier theorem, 152
light-like line, 327

load on a framework, 205
  – equilibrium, 205, 206
  – projective, 210
  – resolvable
    – Euclidean, 206
    – non-Euclidean, 207
Lobachevski, Nikolai Ivanovich, 14, 17, 29, 94
Lobachevsky function, 170
Lorentzian manifold, 324
lune, 4, 11

magnification ratio, 136
matrix of cofactors, 176
Maxwell–Cremona correspondence, 193
  – Euclidean, 217, 220
  – hyperbolic, 226, 227
  – spherical, 222, 224
median, 57, 59, 60, 62
Menelaus of Alexandria, 81
metric plane, 430
  – axioms, 429
  – Bachmann axioms, 428
  – elliptic, 428
  – non-elliptic
    – axioms, 431
Minkowski space, 331, 349
model space, 325

obtuse tetrahedron, 171
octahedron
  – ideal, 170
orthoscheme, 167

Pappus of Alexandria, 68
parabolic space, 339
parallel lines, 29
parallel postulate, 93
pencil of conics, 269
  – double contact pencil, 269, 270, 281
  – dual pencil, 270
Penrose boundary, 359
perfect map (Euler), 125
permanent of a matrix, 171
Pogorelov map, 401
  – infinitesimal, 384
Pogorelov, Alexei Vasilevich, 384
Poincaré disk model of the hyperbolic plane, 160
Poincaré quadratic geometries, 259
polar conic, 268
polarity
  – absolute, 267
polarity (spherical geometry), 20
principles of virtual work, 207
problem
  – Busemann-Petty, 9
  – Castillon, 68
  – Lexell, 28, 42–44
  – Schubert, 47
projective conic, 265
projective distance, 328
projective transformation, 325
pseudo-distance, 326
pseudo-Riemannian manifold, 324
pseudo-sphere, 324
Ptolemy theorem, 153
  – hyperbolic, 153

quasiconformal map, 136

rigidity
  – infinitesimal, 400
rigidity of a framework
  – global, 195
  – infinitesimal, 196
  – local, 195
  – static, 206

Saccheri, Girolamo, 93
Schläfli formula, 177
Schläfli function, 168
Schubert problem, 47
Schubert, Friedrich Theodor von, 47
Seidel conjecture, 171
  – strengthened, 171
Sforza formula, 179
  – compact hyperbolic tetrahedron, 177
  – spherical tetrahedron, 179
shape invariant, 415
  – direct, 415
space-like line, 327
spherical
  – Bretschneider formula, 163
  – Casey theorem, 166
spherical conic, 276
  – center, 276
  – director, 280
  – directrix, 280
  – focal line, 277, 281
  – focus, 278, 281
Study, Eduard, 237, 258–260, 418
support function, 345, 352
support plane, 334
system of forces
  – as a bivector, 209
  – Euclidean, 205
  – non-Euclidean, 206
tetrahedron
  – birectangular, 167
  – Gram matrix, 171
  – ideal, 169
  – obtuse, 171
Theorem
  – Cagnoli, 152
  – Casey, 164
    – hyperbolic, 164
    – spherical, 166
  – Chasles, 272
  – Euler, 3, 8, 18, 28, 152
  – Gauss–Bonnet, 113, 119, 120
  – Girard, 3–5, 7, 8, 13, 43
  – Jacobi, 176
  – Lhuillier, 152
  – Ptolemy (hyperbolic), 153
Theorema egregium, 115, 126
Thurston, William P., 9
Tilly, Joseph-Marie de, 94
time inequality, 255
timelike geometry, 259
timelike line, 327
Tissot, Nicolas-Auguste, 136, 138
trigonometric formulae (de Tilly's proof), 95

volume form, 373

Wallis, John, 7
Weyl formula, 382